Author:
Chambers, Anna Louise

Title:
Transcription termination by a transcription-repair coupling factor

TRANSCRIPTION TERMINATION BY A TRANSCRIPTION-REPAIR COUPLING FACTOR

Anna Louise Chambers

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Ph.D. in the Faculty of Medical and Veterinary Science, Department of Biochemistry, January 2005.

Word Count 70,277

ABSTRACT

Transcription-coupled DNA repair pathways prioritise certain types of damage located in the transcribed strand of active genes for repair. Some lesions in the transcribed strand block the progress of transcription elongation complexes, with the result that the lesion becomes less accessible to repair enzymes. To carry out preferential strand-specific repair, the stalled RNA polymerase (RNAP) must be removed and repair enzymes recruited to the site of damage. These functions are performed by a transcription-repair coupling factor, which in bacteria is a 130 kDa protein called Mfd. Mfd possesses RNAP binding, DNA binding, ATP hydrolysis and UvrA binding activities and is a superfamily 2 helicase that releases stalled elongation complexes in an ATP-dependent manner.

Using site-directed and random mutagenesis, single amino acid substitutions that abolish the ability of Mfd to displace stalled elongation complexes have been identified. In addition a novel *in vivo* assay for Mfd function has been established. The substitutions fall in three regions of Mfd. Firstly, a substitution in the 'Q-tip' motif abrogated the ability to bind and hydrolyse ATP or CTP, required to power elongation complex displacement. Secondly, substitutions within the RNAP interaction domain caused defects in RNAP binding or displacement *in vivo* and *in vitro*. Finally, substitutions have been identified in a region of Mfd that is homologous to the TRG motif of RecG, which disrupt the ability to displace RNAP *in vivo* and *in vitro* but that do not prevent ATP hydrolysis or binding to DNA. These substitutions implicate the TRG motif of Mfd in coupling ATP hydrolysis to translocation along double-stranded DNA and suggest a mechanism for the displacement of transcription elongation complexes.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Nigel Savery for all his helpful advice and interest throughout my three years in the lab. and during the writing of this thesis. I am grateful to Dr. Richard Sessions for his help with molecular modelling and to all the members of D40, past and present, who have given me assistance and made my time in the lab. so enjoyable. Special mentions should go to Abby, Rachel and Tracey not only for providing lots of help, but also for making our bay so friendly and pleasurable to work in. Thanks also go to my other friends, especially Laura and Rich, for their encouragement and for providing lots of fun distractions. Finally, thanks to my family for all their support during my time in Bristol.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree.

Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

SIGNED: A. Chambers DATE: 5/4/05

ABBREVIATIONS

8-oxoG	8-oxo guanine
ApU	Adenylyl (3′→5′) uridine
ATP	Adenosine triphosphate
ATPγS	Adenosine 5-O-(3-thiotriphosphate)
APS	Ammonium peroxodisulphate
BER	Base excision repair
bp	Base pairs
BSA	Bovine serum albumin
B. subtilis	*Bacillus subtilis*
CAT	Chloramphenicol acetyl transferase
CCLR	Cell culture lysis reagent
CIAP	Calf intestinal alkaline phosphatase
CTP	Cytosine triphosphate
E. coli	*Escherichia coli*
DBS	DNA binding site
DHFR	Dihydrofolate reductase
DNA	Deoxyribonucleic acid
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
dNTPs	Deoxynucleoside triphosphates
dsDNA	Double-stranded DNA
DMSO	Dimethylsulphoxide
DTT	Dithiothreitol
EC	Elongation complex
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
HBS	Hybrid binding site

GTP	Guanosine triphosphate
H. pylori	*Helicobacter pylori*
IPTG	Isopropyl-β-D-thiogalactopyranoside
LB	Luria Bertani broth
MBP	Maltose binding protein
MWCO	Molecular weight cut-off
MFD	Mutation frequency decline
NER	Nucleotide excision repair
NTPs	Ribonucleoside triphosphates
OD	Optical density
ONPG	O-nitrophenyl-β-D-galactopyranoside
PCR	Polymerase chain reaction
ppGpp	Guanosine tetraphosphate
PEG	Polyethylene glycol
RBS	RNA binding site
RID	RNA polymerase interaction domain
RNA	Ribonucleic acid
RNAP	RNA polymerase
RT-PCR	Real time polymerase chain reaction
s.d.	Standard deviation
SD	Synthetic dextrose minimal media
SDS	Sodium dodecyl sulphate
SF1	Superfamily 1
SF2	Superfamily 2
ssDNA	Single-stranded DNA
TCR	Transcription-coupled repair
TEMED	N, N, N′-Tetramethylethylenediamine
TRCF	Transcription-repair coupling factor
TRG	Translocation in RecG
UTP	Uridine triphosphate
UV	Ultra-violet
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactoside

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1

INTRODUCTION

In active genes certain types of lesions in the transcribed strand are repaired approximately 10 times faster than in the bulk of the genome (Mellon and Hanawalt, 1989; Mellon *et al.*, 1987), a phenomenon known as transcription-coupled repair. A model for transcription-coupled repair was proposed to explain the non-uniform rates of repair across the genome (Selby and Sancar, 1993a). The model states that some DNA lesions located in the template strand cause stalling of transcription elongation complexes. The presence of an elongation complex stalled at a lesion results in a decrease in the accessibility of that lesion to repair enzymes. Therefore, in order to repair the damage at a rate faster than that of global repair, the stalled elongation complex must be removed and repair enzymes must be recruited to the site. In prokaryotic transcription-coupled repair, RNA polymerase (RNAP) displacement and the recruitment of repair enzymes is carried out by a protein known as a transcription-repair coupling factor (TRCF). Since the first role of the TRCF protein in transcription-coupled repair is the removal of stalled elongation complexes, it is of interest to examine the properties of elongation complexes and to consider pausing of transcription and other mechanisms of termination. This may provide insights into the mechanism of TRCF function.

TRANSCRIPTION

Overview

The process of transcription can be broadly divided into three phases: initiation, elongation and termination (reviewed in (von Hippel *et al.*, 1984). During initiation a promoter sequence is bound by RNAP and is melted to result in an open complex. Several rounds of abortive initiation follow, releasing short transcripts that are

usually two to nine nucleotides long (Hsu *et al.*, 2003). Eventually promoter clearance will occur, accompanied by transition of the transcription complex to a stable, processive ternary elongation complex, which proceeds to transcribe the template at a non-uniform rate, pausing at certain sequences (Landick, 1997). Elongation continues until a termination site is encountered. At termination sites the ternary complex is dismantled and RNAP and the nascent transcript are both released (Nudler and Gottesman, 2002).

The catalytically competent RNAP core enzyme from *Escherichia coli* is composed of five subunits, $\alpha_2\beta\beta'\omega$, that form a large complex of approximately 400 kDa. The core enzyme is bound by another subunit, σ, to form a holoenzyme that is competent in promoter recognition. The σ subunit is responsible for binding to –35 and –10 promoter elements via conserved regions 4.2, 2.3, 2.4 and 2.5 (deHaseth *et al.*, 1998). Following promoter recognition, strand separation occurs between approximately -10 and +1 (the transcription start site) and once an open complex has been formed RNA synthesis can begin. Promoter clearance and entry to the elongation phase occurs after transcription of approximately nine to twelve nucleotides. The majority of work states that upon promoter clearance, the σ subunit is released from the core complex (Gruber *et al.*, 2001). However, two studies using fluorescence techniques have shown that a fraction of elongation complexes retain σ (Bar-Nahum and Nudler, 2001; Mukhopadhyay *et al.*, 2001). Therefore it may be that factors acting on elongation complexes need to be able to do so in the context of bound σ and that σ may have a role in regulation of elongation complexes; for example the recognition of pause sequences at +17 in the bacteriophage λP_R gene, and in *plac* is dependent on σ (Brodolin *et al.*, 2004; Nickels *et al.*, 2002).

Within the elongation complex there is a transcription bubble containing approximately fourteen or fifteen bases of melted DNA. Eight or nine of these melted bases of the template strand form a DNA-RNA hybrid with the newly synthesised RNA transcript. The transcription bubble is stabilised by DNA-protein and nucleic acid interactions within the elongation complex. The elongation complex moves in a 5′ to 3′ direction along the coding strand (3′ to 5′ along the template strand) with RNA being synthesised 5′ to 3′ as NTPs are incorporated at the active site.

Termination of transcription is characterised by collapse of the transcription bubble and release of RNAP and RNA from the DNA template. The RNAP core that is released is then able to bind another σ subunit and reinitiate transcription.

The structure of transcription complexes

Elongation complexes are capable of incorporating 60-80 nucleotides per second into nascent RNA (Gotta *et al.*, 1991; Vogel and Jensen, 1994), yet at the same time are extremely stable and processive. These properties are essential for the production of long transcripts, since following termination RNA synthesis cannot be restarted without reinitiation from the promoter. Whilst open complexes fall apart under high salt conditions, elongation complexes are salt stable. How is this stability achieved? The answer appears to be a combination of nucleic acid and protein-nucleic acid interactions.

Crystal structures of both the holoenzyme (Vassylyev *et al.*, 2002) and core enzyme (Zhang *et al.*, 1999) have been determined at 2.6 Å and 3.3 Å respectively. The structure of RNAP holoenzyme from *Thermus thermophilus* shows that the enzyme is shaped like a crab claw with the β subunit forming one 'pincer' and the β′ subunit

the other. Between these two jaws there is a central channel about 27 Å wide running the length of the protein. The active centre is located at the base of the channel where the β and β′ subunits interact and a Mg^{2+} ion is chelated at the active site by the NADFDGD motif of β′. The β and β′ subunits contain several evolutionarily conserved colinear sequences (A-I in β and A-H in β′) (Jokerst *et al.*, 1989; Sweetser *et al.*, 1987), which are clustered around the active site in the crystal structures. The conserved β′F region forms a bridging helix that traverses the main channel opposite the active site. In combination with the β′G region, the β′F helix splits the main channel in two, creating a smaller secondary channel that is unable to accommodate double stranded DNA. This secondary channel permits nucleotide entry to the active site and is also the channel through which the 3′ end of RNA is extruded in backtracked elongation complexes. Other features of note include the β′C region, which forms a coiled-coil rudder that extends into the main channel and the βG region, which forms a flap that is connected to the core enzyme by a flexible linker. Since the DNA backbone remains intact during initiation, open complex formation must involve the opening of the jaws of RNAP to allow the nucleic acid to enter the main channel.

The σ subunit is composed of four structural domains: two N-terminal domains, a linker domain and a C-terminal domain, and is located primarily on the surface of the core enzyme. Although several regions of σ are involved in creating a large interaction area with the core enzyme, the most extensive contacts are formed between the first N-terminal domain of σ and the β′ subunit. The two N-terminal domains of σ form a V-shaped structure located close to the aperture of the channel for downstream DNA. This section of σ contains regions 2.3-2.5 and these are

responsible for binding to the –10 and extended –10 regions of the promoter (Barne *et al.*, 1997; deHaseth *et al.*, 1998). The C-terminal domain binds to the –35 promoter element. A short region of the linker domain of σ is inserted into the active site channel and extends through the RNA exit channel. Synthesis of a twelve nucleotide transcript would result in a clash with this domain; therefore the presence of nascent RNA may aid the dissociation of σ upon entry to elongation.

During the transition into elongation, the conformation of RNAP alters dramatically. The structure of the core enzyme from *Thermus aquaticus* shows that RNAP retains the same basic crab claw shape as the holoenzyme (Zhang *et al.*, 1999) (figure 1.1 A), although the main channel is more open than in the initiation complex structure. The most significant changes in conformation occur within the βG flap domain, the αI subunit, the N-terminal section of β and the coiled-coil domain of β′. These changes result in the RNA exit channel being most affected upon promoter clearance.

Based on crosslinking and mapping data (Nudler *et al.*, 1998), DNA and RNA were modelled into the core structure (Darst, 2001). In this model the angle between the DNA entering the enzyme (downstream) and the DNA exiting the enzyme (upstream) is around 90°. The major contribution to the stability of the elongation complex is from interactions between the nucleic acid scaffold and RNAP. Three major nucleic acid-protein interaction sites have been characterised in the elongation complex: the double-stranded DNA binding site, the DNA:RNA hybrid binding site and a binding site for the single-stranded RNA transcript (figure 1.1 B). The downstream DNA binding site is formed by a trough within the N-terminus of β′, which is enclosed by the C-terminal region of β. The interaction with dsDNA was shown to be non-ionic, based on its resistance to high salt, and

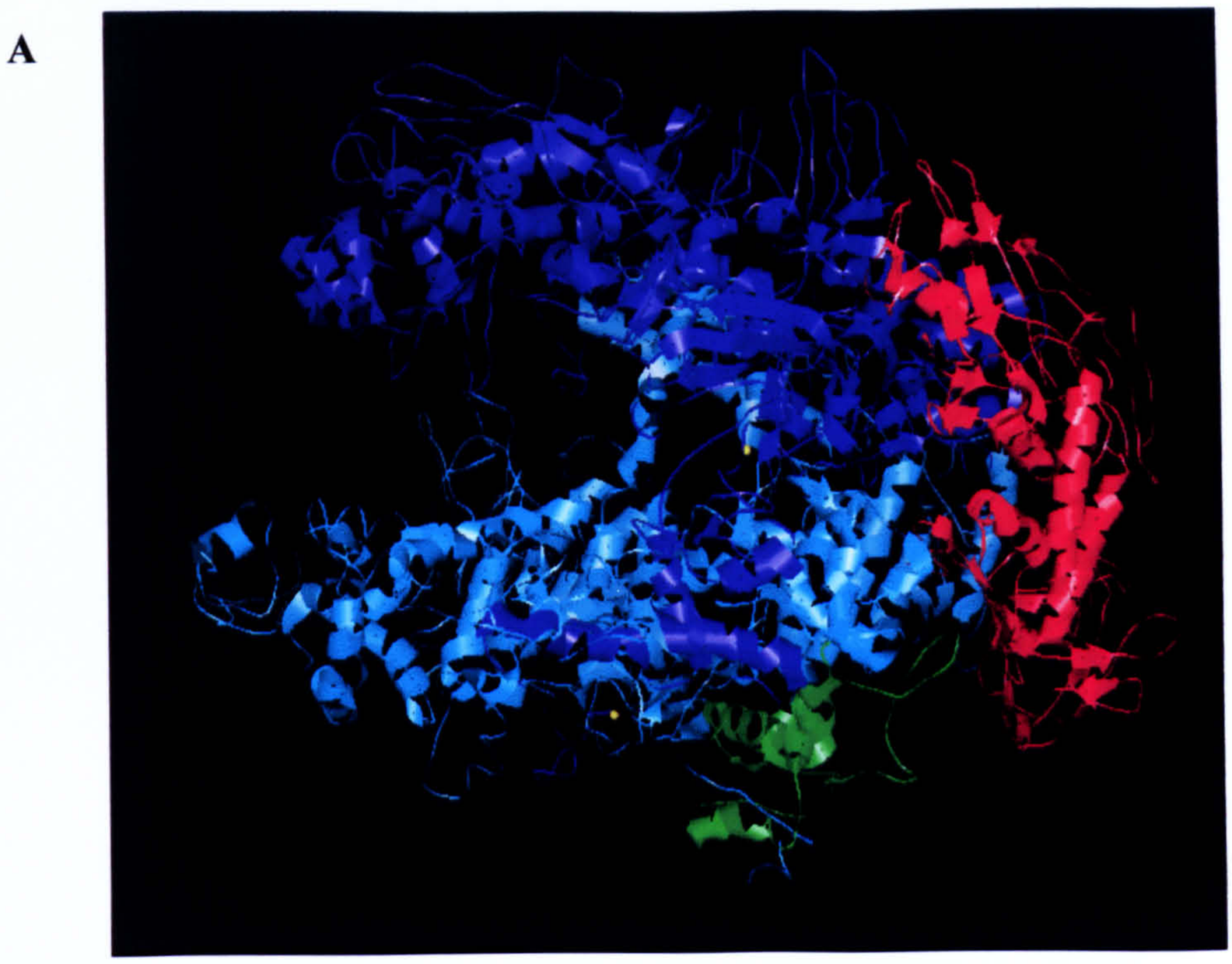


Figure 1.1. The Transcription Elongation Complex

(A) The structure of core RNAP from *Thermus aquaticus* (Zhang *et al.*, 1999). The α subunits are shown in red, the ω subunit in green, β in blue and β′ in turquoise. In the view shown the upstream DNA would come out of the page.

(B) A model of the transcription bubble and EC based on crosslinking and footprinting DNA adapted from (Darst, 2001; Komissarova and Kashlev, 1998; Zaychikov *et al.*, 1995). The template strand is shown in maroon, the non-transcribed strand in lilac, the transcript in purple and the nucleotide being incorporated at the active site in green. Open purple circles indicate the location of RNA in backtracked complexes. The grey box shows the protection of the nucleic acid scaffold by RNAP from footprinting studies. RBS, HBS and DBS correspond to the RNA-binding site, the hybrid-binding site and DNA-binding site, respectively.

requires 9 bp (base pairs) of duplex DNA downstream of the 3′ end of the RNA (Nudler *et al.*, 1996). The non-template strand is located between two regions of the β subunit and the interaction with the nucleic acid may be tightened by a change in conformation of the enzyme, such as movement of the bridging helix to permit closure of the main channel.

The RNA:DNA hybrid is accommodated by the main channel and substitutions that confer rifampicin-resistance are found in parts of the β subunit that line this region (Campbell *et al.*, 2001). Rifampicin has been proposed to prevent entry into the elongation phase by blocking the path of RNA after synthesis of only two or three nucleotides. The hybrid-binding site relies on ionic interactions between the protein and approximately six nucleotides immediately upstream of the active site (Nudler *et al.*, 1996). Crosslinking experiments have shown that the template DNA remains in register with the nascent RNA, forming a hybrid for 8-9 bp (Nudler *et al.*, 1997). Reconstitution of synthetic ECs has shown that the hybrid makes a contribution to the stability of the complex (Wilson *et al.*, 1999) and this has also been shown to be the case for natural ECs. Important contacts also seem to be formed between the protein and the RNA in the region where the transcript is separated from the hybrid. The β′C region forms a coiled-coil rudder that extends into the main channel near where the template strand and hybrid separate and the β′ rudder region crosslinks to probes incorporated at –10 in the RNA.

The RNA binding site was localised to the region immediately upstream of the hybrid binding site, consistent with RNase protection data (Komissarova and Kashlev, 1997). After being stripped off the hybrid by the rudder, the nascent transcript exits through a channel beneath the βG flap domain. The RNA-binding site is comprised of the N-terminus of β′ and the C-terminus of β, the same regions

that are involved in the DNA binding site, indicating that the DNA entry and RNA exit channels are on the same face of the protein.

RNAP appears to be responsible for maintenance of hybrid length and the length of the transcription bubble as well as providing important contacts with the nucleic acid scaffold that provide stability. In order to terminate transcription, the stability and structure of the elongation complex needs to be disrupted and this could be accomplished by weakening interactions within either the DNA-binding site, the hybrid-binding site or the RNA-binding site or by affecting subunit interactions within RNAP.

Stalling of transcription elongation

Transcription elongation does not proceed at a uniform rate; certain sequences in the template cause the elongation complex to pause or even arrest (Landick, 1997). Naturally occurring pause sites cause ECs to temporarily stall but these complexes are able to restart transcription spontaneously. However, during nucleotide starvation or upon blockage of transcription, elongation can become halted and transcription cannot resume without the assistance of extra factors. Elongation complexes that have entered this state are termed arrested complexes. During arrest, RNAP moves backwards on the DNA template and the 3′ end of the RNA loses contact with the active site resulting in backtracking of the elongation complex. As the RNAP becomes backtracked, the 3′ end of the transcript is extruded through the secondary channel, the site for nucleotide entry during elongation (Komissarova and Kashlev, 1997). The secondary channel has recently been identified as being the binding site for both a regulatory protein called DksA that increases sensitivity to ppGpp (Paul *et al.*, 2004; Perederina *et al.*, 2004) and an antibiotic called microcin

J25 (Adelman *et al.*, 2004; Mukhopadhyay *et al.*, 2004). Both of these molecules insert into the secondary channel and are able to prevent backtracking of RNAP.

In vivo, pausing and arrest may be diminished compared to *in vitro* single-round experiments that are commonly used to study transcription. Under multi-round transcription conditions the presence of trailing ECs reduced pausing and arrest of a leading EC (Epshtein and Nudler, 2003). This cooperation between RNAP molecules is likely to reduce transcriptional delays on highly transcribed templates *in vivo,* with multiple elongation complexes pushing forwards any complexes in front of them that have halted at pause sites or protein roadblocks, preventing them entering backtracked states.

Pausing and arrest of transcription is controlled in part by proteins that interact with the elongation complex. There are a number of proteins that are able to bind to RNAP to regulate the frequency of pausing. These include the Nus proteins, which also affect termination efficiency and are therefore discussed in the section on termination. There are also proteins that are able to act on backtracked complexes to return them to productive elongation.

Proteins that act on arrested elongation complexes - GreA and GreB

GreA and GreB are proteins that permit the reactivation of arrested backtracked elongation complexes. Both proteins stimulate an intrinsic endonuclease cleavage activity of RNAP to remove the 3′ end of the RNA transcript that has become detached from the active site (Laptenko *et al.*, 2003). The realignment of the 3′ end of the transcript with the active site allows RNA synthesis to resume. GreA is responsible for cleavage of short di- and tri-nucleotides of RNA while GreB causes cleavage of longer regions of RNA. The C-terminal domains of both GreA and GreB bind near the opening of the secondary channel of RNAP (Laptenko *et al.*,

2003), through which nucleotides access the active site and through which RNA is extruded during backtracking. The N-terminal domain of both GreA and GreB appears to insert into the secondary channel and donate two acidic amino acids, located at the end of a coiled-coil, to co-ordinate a magnesium ion that is usually weakly bound at the active site of RNAP and as a result RNA cleavage is stimulated. A low resolution cryo-EM structure (Opalka *et al.*, 2003) combined with docking of the high resolution structure of GreB with the RNAP core structure suggests that there is room for both Gre and extruded RNA within the secondary channel.

Termination of transcription

Terminators are defined as sites at which release of RNA is more probable than incorporation of the next nucleotide. Prokaryotes utilise two classes of termination signal to dissociate the highly stable EC, both of which act through the nascent transcript. Intrinsic terminators are composed of a GC-rich stem, which is capable of forming a hairpin, followed by a run of U residues whilst Rho-dependent terminators require binding of the Rho protein to a *rut* site within the nascent transcript. Both termination mechanisms involve a sequence-specific pause of transcription followed by destabilisation of the paused complex, resulting in release of both RNAP and the transcript from the DNA. During transcription-coupled repair the TRCF protein causes displacement of a stalled elongation complex and therefore it is of interest to consider the mechanisms of elongation complex destabilisation at intrinsic and Rho-dependent termination sites in more detail.

Intrinsic termination

Intrinsic terminators are short sequences containing a region that encodes a hairpin in the nascent RNA, followed immediately by a run of 5-9 U residues just before the

site of termination. Initially it was proposed that the presence of a weak dA:rU hybrid at a terminator was solely responsible for spontaneous RNAP dissociation due to a decrease in EC stability (von Hippel, 1998; Wilson *et al.*, 1999). An alternative model suggested that formation of the hairpin decreased interactions between the RNA binding site of RNA polymerase and the RNA in order to destabilise the EC (Gusarov and Nudler, 1999). A variation of this model proposed that hairpin formation pulls the RNA out of the transcription complex (Yarnell and Roberts, 1999). This could occur without movement of RNAP on the DNA template, or alternatively the RNAP and transcription bubble could translocate forwards without the concurrent synthesis of RNA.

Single molecule studies have observed pausing prior to termination (Yin *et al.*, 1999) and deletion of the hairpin region of a terminator but not the run of U residues caused the EC to stall at the site where termination normally occurs. This indicates that the stretch of U residues is responsible for stalling the transcription complexes (Komissarova *et al.*, 2002) in order to permit folding of the hairpin close to the RNAP. Following pausing of the transcription complex, formation of the RNA hairpin is crucial for destabilisation of the elongation complex. As stated above, it has been shown that an EC containing a transcript that possesses only the U tract of the terminator is more stable than one containing the U tract and a hairpin, establishing the importance of hairpin formation in termination (Komissarova *et al.*, 2002). Indeed, annealing a DNA oligonucleotide to the RNA to substitute for the upstream arm of the hairpin also results in termination (Yarnell and Roberts, 1999). Furthermore it was shown by permanganate footprinting and by increasing the stability of the hybrid immediately downstream of the hairpin stem, that hairpin folding disrupts the 8 bp hybrid, shortening it to 5 bp which decreases the stability of

the EC (Komissarova *et al.*, 2002). The folding of a hairpin is energetically favourable and would be expected to compensate for the cost of breaking the rU:dA hybrid. The formation of the hairpin may also reduce stability of the transcription complex by removing interactions between the nascent RNA and the RNA exit channel. Overall these studies support the model that during pausing induced by the stretch of U residues, the nascent RNA folds into a hairpin and that hairpin formation extracts the nascent RNA from the transcription complex. This has the result of shortening the hybrid and destabilising the EC by disrupting interactions within the hybrid- and RNA-binding sites.

A variation of this model, that upon hairpin formation the transcription bubble and RNAP translocate forward without RNA synthesis, is supported by a number of pieces of evidence. Proteins bound to the DNA immediately downstream of the terminator prevent termination, presumably by preventing forward movement of RNAP (Nudler *et al.*, 1995). If rewinding of the template strand with the non-template strand at the upstream end of the transcription bubble is prevented, termination is impaired (Ryder and Roberts, 2003), indicating that rewinding of upstream DNA and unwinding of the hybrid is involved in termination. The forward translocation model of termination was also examined by preventing melting of downstream DNA and RNAP forward movement (Santangelo and Roberts, 2004). When interstrand psoralen crosslinks were introduced downstream of an intrinsic terminator to prevent melting, the efficiency of termination was reduced but was not completely abolished. When a roadblock protein was bound immediately downstream of the elongation complex, and an oligonucleotide was annealed to the RNA to form a hairpin, termination was blocked. Termination was restored if the protein roadblock was moved 4 bp downstream to permit forward movement of

RNAP. This is consistent with the fact that a reduction in the length of the hybrid from 8-9 bp to 5 bp causes elongation complex destabilisation. The authors proposed that the usual mechanism of termination involves forward movement of RNAP and the associated bubble, but when translocation of the bubble is inhibited, termination can occur via a different mechanism, which may involve translocation of RNAP on double-stranded downstream DNA.

The structure of core RNAP shows that there is insufficient room under the βG flap of the RNA exit channel for a hairpin. Therefore hairpin formation may cause a conformational change that reduces the stability of the elongation complex (Toulokhonov *et al.*, 2001), although if the RNAP translocates forwards it is feasible that hairpin formation could occur outside of the RNA polymerase. Recently, mutation of the conserved zinc-binding domain near the N-terminus of β has been shown to decrease termination efficiency (King *et al.*, 2004) and it is proposed that this domain may interact with terminator RNA to promote hairpin folding or pausing.

Rho-dependent termination

During Rho-dependent termination the Rho protein binds to a *rut* site in the nascent transcript and acts on the RNA to bring about ATP-dependent RNAP dissociation (Richardson, 1996). Like intrinsic termination, Rho-termination requires a decreased rate of elongation by RNAP brought about by a pause site in order to cause termination. Rho is a hexameric ring protein with each subunit containing a single-stranded RNA binding domain and an ATPase domain (Bogden *et al.*, 1999; Kaplan and O'Donnell, 2003). The RNA appears to bind around all six subunits, but only three of the subunits may bind RNA with high affinity simultaneously (Geiselmann *et al.*, 1992a). Similarly only three of the subunits can simultaneously bind ATP

(Stitt, 1988). Once Rho binds to RNA, ATP hydrolysis powers translocation of Rho along the RNA until it reaches the RNAP where it brings about dissociation of the transcription complex (Geiselmann *et al.*, 1992b). Since Rho acts on the RNA transcript it is likely to cause termination either by disrupting the interaction between the nascent transcript and the RNA binding site of RNAP or by extracting the RNA from the EC, or a combination of both. The presence of ribosomes on the RNA prevents binding of Rho and therefore inhibits Rho-dependent termination. This provides a means to increase the specificity of termination since the RNA will be most accessible within non-coding regions of the mRNA such as at the end of an operon.

Proteins that affect termination efficiency - NusA, NusG, Q and N

The efficiency of termination is not simply determined by terminator sequences in the DNA or RNA, or by the Rho protein. There are a number of proteins that can regulate the relative efficiency of read-through or termination at these sequences including the Nus proteins, which are also involved in regulation of pausing. In general these proteins act by stabilising or destabilising the interaction between the RNAP and the transcription bubble, in particular interactions within the RNA-binding site. NusA binds to transcription elongation complexes and increases the efficiency of intrinsic termination. NusA brings about an increase in pausing and destabilises the interaction between single-stranded RNA and RNAP to favour hairpin formation and therefore termination. NusA can only bind to RNAP once σ has dissociated i.e. following promoter clearance. The structure of NusA (Shin *et al.*, 2003) shows a globular N-terminal head domain, which possesses several negatively charged side chains that could potentially interact with the positively charged βG flap of RNAP, the flap under which the RNA exits the elongation

complex. The protein also contains three RNA binding domains but NusA on its own cannot bind RNA *in vitro*, requiring interaction with RNAP to bind RNA, preferentially at a *nut* site.

N is an antitermination protein from phage λ that counteracts the effect of NusA in early operons of bacteriophage λ; in fact NusA stimulates antitermination by N. N decreases pausing although not to an extent that accounts for the suppression of termination that is observed. The mechanism of antitermination by N requires specific protein-protein and protein-RNA interactions between N and the EC. The actions of N and NusA have been extensively examined (Gusarov and Nudler, 2001). When N was provided at a high enough concentration antitermination occurred even in the absence of NusA and a *nut* site, indicating that the function of the *nut* site is to increase the local concentration of N by recruiting it to the EC. Like NusA, N does not affect the stability of the DNA:RNA hybrid but predominantly affects interactions in the RNA binding site. By incorporation of a photocrosslink into the RNA, NusA and N were both shown to contact the nascent RNA approximately 18-24 nucleotides from the catalytic centre. This region of the RNA transcript corresponds to the upstream half of the termination hairpin, implying that NusA and N control termination by their effects on hairpin formation. NusA binding resulted in weakening of the interaction between RNAP and RNA without making any strong compensatory interactions. On the other hand, N did not decrease crosslinking between RNAP and the RNA and also provided a further interaction with the shoulder of the hairpin, with the overall effect of strengthening the affinity of RNAP for RNA. Upon interaction with N, the RNA binding domains of NusA are exposed to further increase the affinity of the EC for RNA.

λQ is another antitermination protein but it specifically acts on transcription complexes that have initiated from the P_R promoter of λ. Like N, NusA enhances the antitermination properties of Q. A pause (*qut*) just downstream of the promoter permits Q to interact with RNAP and the resulting polymerase is capable of transcribing through both intrinsic and Rho-dependent terminators. In addition complexes that are modified by Q do not remain at pause sites as long as unmodified RNAP. The antitermination effect of Q is not solely due to an increase in transcription rate, which could result in transcription complexes reading through termination sites. Termination can be induced by the addition of oligonucleotides that anneal to RNA to mimic the formation of a termination hairpin (Yarnell and Roberts, 1999). Termination by this mechanism was abolished when static complexes were modified with Q, implying that Q increases the stability of the transcription complex (Yarnell and Roberts, 1999). Therefore *in vivo* Q may act to both decrease the frequency of pausing that precedes termination and to inhibit hairpin formation.

NusG is an essential protein that increases Rho-dependent termination. NusG interacts directly with both RNAP and with Rho (Sullivan and Gottesman, 1992) and is able to increase the rate of transcription elongation (Burova *et al.*, 1995). It is also required for maintenance of the high rate of elongation during *rrn* antitermination (Zellars and Squires, 1999) and is essential for N antitermination (Li *et al.*, 1992). NusG brings about an increased rate of transcription by suppression of pause sites. It has been suggested that the rate of elongation and antitermination are closely linked and therefore the accelerated rate of transcription in the presence of NusG may be responsible for its antitermination effects.

A protein that terminates transcription independently of a terminator - T4 Alc

The Alc protein from bacteriophage T4 terminates *E. coli* RNAP transcription prematurely only if the transcription complex is moving fast (Kashlev *et al.*, 1993), contrasting noticeably with intrinsic and Rho-dependent termination in which the EC is paused prior to its displacement from the template. In addition termination by Alc does not require a termination sequence. Alc only terminates transcription on DNA that contains cytosine and therefore inhibits host transcription but not transcription of its own DNA, which contains modified cytidine bases (Snyder *et al.*, 1976). It was shown that a dispensable region (~ amino acids 165-328) of the β subunit of RNAP located between conserved motifs B and C was required for the interaction with, and function of, Alc (Severinov *et al.*, 1994). Mutations in this region of RNA polymerase III from yeast have also been shown to affect termination, perhaps linking this region of RNAP to the binding of some termination factors.

DNA DAMAGE AND GLOBAL REPAIR

Types of DNA damage and repair pathways

DNA is constantly subject to attack by both endogenous and exogenous mutagens that cause a variety of DNA damage. In order to maintain the integrity of the genetic material, cells possess a number of repair pathways to recognise and repair these different types of damage. Mismatches, deletions or insertions may arise during replication of the DNA and these are recognised by the mismatch repair pathway, which in prokaryotes involves the MutS, MutL and MutH proteins (reviewed in (Yang, 2000). Alternatively oxidative damage can cause bases to be modified or lost and these altered or abasic sites are recognised by a slew of base excision repair

(BER) enzymes that remove the mutated base before an AP lyase enzyme cuts the DNA backbone (reviewed in (Dizdaroglu, 2003). In addition, ionising radiation can cause the DNA backbone to break or strands to become cross-linked.

On exposure to UV irradiation neighbouring pyrimidine bases can become cross-linked to form pyrimidine dimers and 6-4 photoproducts. These DNA lesions can be repaired by a photoreactivating enzyme called photolyase, which upon recognition of bulky helix-distorting lesions, absorbs near-UV and blue light. The absorption of a photon causes excitation of an electron that is able to directly cleave the cross-links between the two pyrimidine bases (reviewed in (Carell *et al.*, 2001). However, the major pathway for repair of bulky lesions, including thymine dimers, is nucleotide excision repair (NER).

Nucleotide Excision Repair in prokaryotes

The prokaryotic NER pathway contains the UvrA, B and C proteins (reviewed in (Friedberg *et al.*, 1995; Van Houten, 1990). UvrA acts as the damage recognition protein and binds to UvrB to form a $UvrA_2B$ complex that delivers UvrB to the site of damage. UvrA dissociates to leave a UvrB-DNA complex that is bound by UvrC. UvrC acts as an endonuclease to cleave the fifth phosphodiester bond 3′ to the lesion followed by the eighth phosphodiester bond on the 5′ side. Following excision of an oligomer containing the lesion, UvrD acts as a helicase to remove the excised 12 or 13-mer leaving DNA polymerase I to fill in the gap, using the intact strand as a template. Nicks in the backbone are sealed by DNA ligase to complete repair (figure 1.2).

The 104 kDa (940 amino acid) DNA-binding protein UvrA is required for DNA damage recognition during NER. Deletion constructs have demonstrated that the

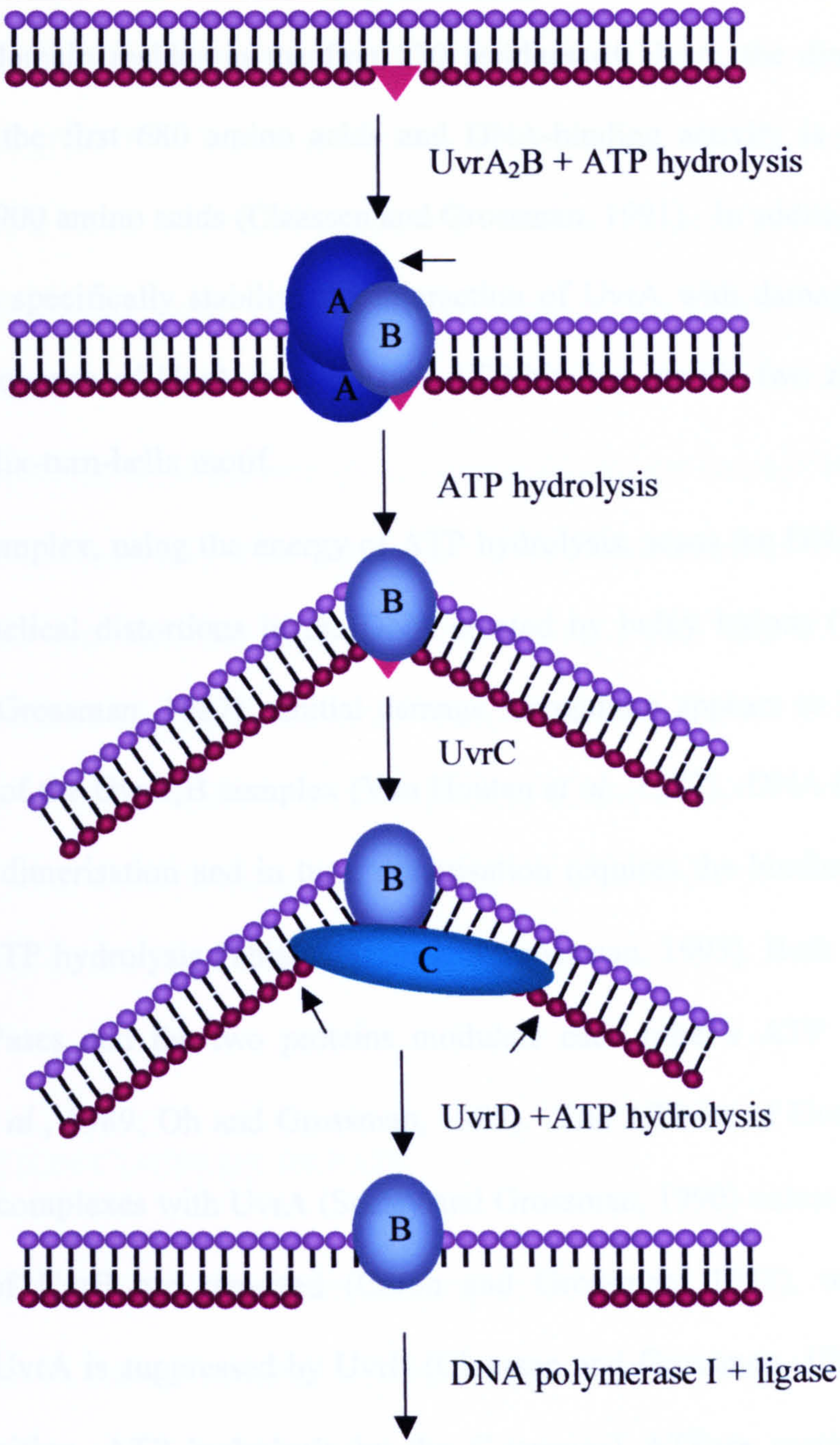


Figure 1.2. Nucleotide Excision Repair.

The $UvrA_2B$ complex recognises bulky lesions in an ATP-dependent manner. On locating damage UvrA dissociates, leaving UvrB bound to the DNA. UvrB is bound by the endonuclease UvrC that nicks the DNA backbone firstly 5 bases away on the 3′ side and then 8 bases away on the 5′ side of the lesion. The oligonucleotide is removed by UvrD before DNA polymerase I fills in the gap using the intact strand as a template. The nicks are sealed by DNA ligase.

UvrB binding domain resides in the first 230 residues of UvrA, the dimerisation domain within the first 680 amino acids and DNA-binding activity is contained within the first 900 amino acids (Claassen and Grossman, 1991). In addition the last 40 amino acids specifically stabilise the interaction of UvrA with damaged DNA. The primary sequence of UvrA includes two ATP-binding motifs, two zinc finger motifs and a helix-turn-helix motif.

The $UvrA_2B$ complex, using the energy of ATP hydrolysis, scans the DNA 5′ to 3′, searching for helical distortions in the DNA created by bulky lesions (Oh *et al.*, 1989; Oh and Grossman, 1987). Initial damage recognition appears to be via the UvrA subunits of the $UvrA_2B$ complex (Van Houten *et al.*, 1987). DNA binding by UvrA requires dimerisation and in turn, dimerisation requires the binding of ATP, although not ATP hydrolysis (Thiagalingam and Grossman, 1993). Both UvrA and UvrB are ATPases and the two proteins modulate each other's ATP hydrolysis activity (Oh *et al.*, 1989; Oh and Grossman, 1987). The ATPase of UvrB remains cryptic until it complexes with UvrA (Seeley and Grossman, 1990) unless the last 40 amino acids of UvrB are removed (Caron and Grossman, 1988), whilst ATP hydrolysis by UvrA is suppressed by UvrB (Claassen and Grossman, 1991). Upon damage recognition, ATP hydrolysis by the C-terminal ATPase motif of UvrA causes its dissociation from DNA (Thiagalingam and Grossman, 1993), leaving UvrB loaded at the lesion. ATP hydrolysis by UvrB then results in melting of the DNA (Theis *et al.*, 2000).

Crystal structures of UvrB from *Thermus thermophilus* (Machius *et al.*, 1999) and *Bacillus caldotenax* (Theis *et al.*, 1999) show that it is composed of five domains. Two of these domains, 1a and 3, contain the helicase motifs, domain 2 contains the region proposed to interact with UvrA and domain 4 interacts with both UvrA and

UvrC. A change in domain orientation upon UvrA dissociation is anticipated to cause a β-hairpin of UvrB to insert between the two strands of the DNA and lead to the formation of a stable UvrB-DNA complex. Both the tip and the base of the β-hairpin contain several conserved hydrophobic amino acids suitable for stacking with the DNA bases.

A study using scanning force microscopy has proposed that the damage recognition complex has the stoichiometry A_2B_2 rather than A_2B (Verhoeven *et al.*, 2002). The authors postulate that on initial binding, DNA wraps around one of the UvrB subunits of the A_2B_2 complex to search for damage before the other UvrB subunit wraps the other strand of DNA to permit scanning of this strand. Upon damage recognition the UvrB dimer is left on the DNA and UvrC can bind, possibly resulting in the dissociation of one of the UvrB subunits.

TRANSCRIPTION-COUPLED REPAIR

The effect of DNA lesions on transcription elongation

Some types of DNA damage interfere with the translocation of elongation complexes along the DNA template while other lesions, despite resulting in a change in the sequence of the RNA transcript, do not impede progression of transcription (reviewed in (Tornaletti and Hanawalt, 1999). In *E. coli,* the length of nascent RNA decreases with increasing UV dosage (Michalke and Bremer, 1969), suggesting that UV lesions such as pyrimidine dimers and 6-4 photoproducts block transcription. Subsequently it was confirmed *in vitro* that transcription by *E. coli* RNAP was arrested by thymine dimers located in the transcribed strand of the template but not in the non-transcribed strand (Selby and Sancar, 1993a). Similarly, a psoralen

adduct in the coding strand did not affect transcription by *E. coli* RNAP but when placed in the template strand it resulted in arrest of elongation complexes and a DNaseI footprint showed that stably stalled elongation complexes were present at the lesion site (Shi *et al.*, 1987).

In contrast to the lesions described above, 8-oxoG lesions (Viswanathan and Doetsch, 1998) and remarkably, gaps (Zhou and Doetsch, 1993) and abasic sites (Liu and Doetsch, 1996) are either by-passed by *E. coli* RNAP or only transiently inhibit transcription. Those lesions that cause arrest of elongation complexes are in general bulky lesions that are repaired by the nucleotide excision repair pathway and correlate with the types of lesion that are subject to transcription-coupled repair. However there are claims that 8-oxoG lesions are also subject to transcription-coupled repair (Bregeon *et al.*, 2003).

The discovery that RNAP becomes stably halted at lesions in the template strand makes clear the importance of the TCR pathway. Failure to repair lesions in transcriptionally active genes will not only result in the loss of that particular protein but may also affect global transcription due to sequestration of RNAP at sites of damage.

The discovery of TCR and MFD

Unusually, the phenomenon of transcription-coupled repair was first discovered in eukaryotic cells when removal of UV-induced pyrimidine dimers from the DHFR (dihydrofolate reductase) gene of Chinese hamster ovary cells was observed to occur faster than from the overall genome (Bohr *et al.*, 1985). This was the first indication that not all regions of the genome are repaired at a uniform rate, with active genes being preferentially repaired. On closer examination, it was shown that it was the strand that acted as the template for RNA synthesis that was selectively repaired

(Mellon *et al.*, 1987). Transcription-coupled repair was shown not to be specific to mammalian cells; preferential repair of the transcribed strand of the lactose operon in *E. coli* was observed on induction of transcription (Mellon and Hanawalt, 1989) and transcription-dependent strand-specific repair was also seen in yeast (Sweder and Hanawalt, 1992). It was subsequently found that, in bacteria, transcription-coupled repair (TCR) is restricted to sequences a minimum of 15 bases downstream of promoter start points (Selby and Sancar, 1995b).

MFD (mutation frequency decline) is a phenomenon that was first described in 1966 (Witkin, 1966). Following UV irradiation, cells that were placed in a medium that does not support protein synthesis showed a decrease in the frequency of certain tRNA suppressor mutations. Analysis showed that only lesions in the transcribed strand were subject to MFD (Bockrath and Palmer, 1977) and therefore MFD was revealed to be a strand-specific repair phenomenon. Under MFD conditions (amino acid starvation) the stringent response inhibits tRNA synthesis; however MFD occurred at the same rate in a *rel*$^-$ strain with a non-functional stringent response and therefore under conditions where tRNA synthesis was not repressed. Despite the lack of correlation between transcription activity and MFD, MFD resulted in preferential repair of the transcribed strand. A strain which did not undergo MFD was isolated and the gene that was defective was termed the *mfd* gene (Witkin, 1966). The *mfd*$^-$ cells possessed only moderate UV-sensitivity (Selby and Sancar, 1993a).

Mechanism of Transcription-Coupled Repair

How does transcription of a damaged template result in an increased rate of NER of the transcribed strand? Arrest of transcription at a lesion in the template strand somehow acts as a signal to promote repair at that site. One model is that the stalled

transcription complex is directly recognised by the UvrABC repair proteins, leading to their delivery to the damage site, with RNAP effectively acting as a damage sensor. It was also proposed that UvrA is recruited, through interaction with the β subunit of RNAP, to the open DNA at the promoter and then translocates in a 5′-3′ direction along the non-transcribed strand, searching for damage within the template strand (Ahn and Grossman, 1996).

However when the UvrABC proteins were added to an *in vitro* system containing RNAP stalled at a thymine dimer in the template strand, it was the non-transcribed strand that was repaired faster (Selby and Sancar, 1990). This is the reverse of the observation *in vivo* where UV damage in the template strand of transcriptionally active genes is repaired fastest. This indicates that the presence of a stalled elongation complex at a lesion inhibits its repair, rather than promoting it, potentially by impeding access of repair enzymes to the lesion. By implication there is a factor *in vivo* that removes the RNAP stalled at the lesion in order to permit repair and this putative factor was termed a transcription-repair coupling factor (TRCF).

Removal of the stalled elongation complex would abolish inhibition of repair but would be expected to result in a rate of template strand repair no greater than that of the coding strand. Therefore the repair enzymes must be targeted to the lesion to cause the favoured repair of the template strand. Thus the putative transcription-repair coupling factor (TRCF) must possess a second function i.e. the recruitment of repair enzymes, in addition to the removal of stalled RNAP from DNA damage sites. A model of TCR was proposed in which a lesion in the template strand blocks transcription by RNAP (Selby and Sancar, 1993a). This stalled elongation complex is recognised by a TRCF that causes release of RNAP and the nascent transcript. The TRCF may replace the transcription complex at the damage

site and recruit the $UvrA_2B$ complex to perform nucleotide excision repair (figure 1.3).

An *in vitro* system composed of purified *E. coli* RNAP, UvrA, UvrB, UvrC, UvrD, DNA polymerase I and ligase did not carry out strand-specific repair of a UV irradiated template despite being proficient at nucleotide excision repair (Selby and Sancar, 1990). However preferential repair of the transcribed strand was observed when the system was supplemented with certain fractions of an *E. coli* cell extract, confirming the existence of a transcription-repair coupling factor. Using this system the 130 kDa TRCF of *E. coli* was partially purified (Selby and Sancar, 1991).

The Mfd protein is the bacterial TRCF

When *in vitro* transcription-coupled repair assays indicated the presence of a TRCF protein in cell extract, the product of the *mfd* gene was considered as a possible candidate due to its requirement for the preferential repair of the transcribed strand of tRNA genes during MFD (Bockrath and Palmer, 1977). *E. coli* cell extracts from a wild-type and *mfd*⁻ strain were added to an *in vitro* assay for strand-specific repair (Selby *et al.*, 1991). In the absence of transcription the rate of repair was the same for both strands with either cell extract. Under conditions where transcription could occur, the addition of wild-type cell extract brought about preferential repair of the transcribed strand, but this did not occur using the extract from *mfd*⁻ cells indicating that the Mfd protein is the bacterial TRCF. This was further supported by the fact that when *mfd*⁻ cell extract was supplemented with partially purified TRCF, preferential repair was re-established (Selby *et al.*, 1991).

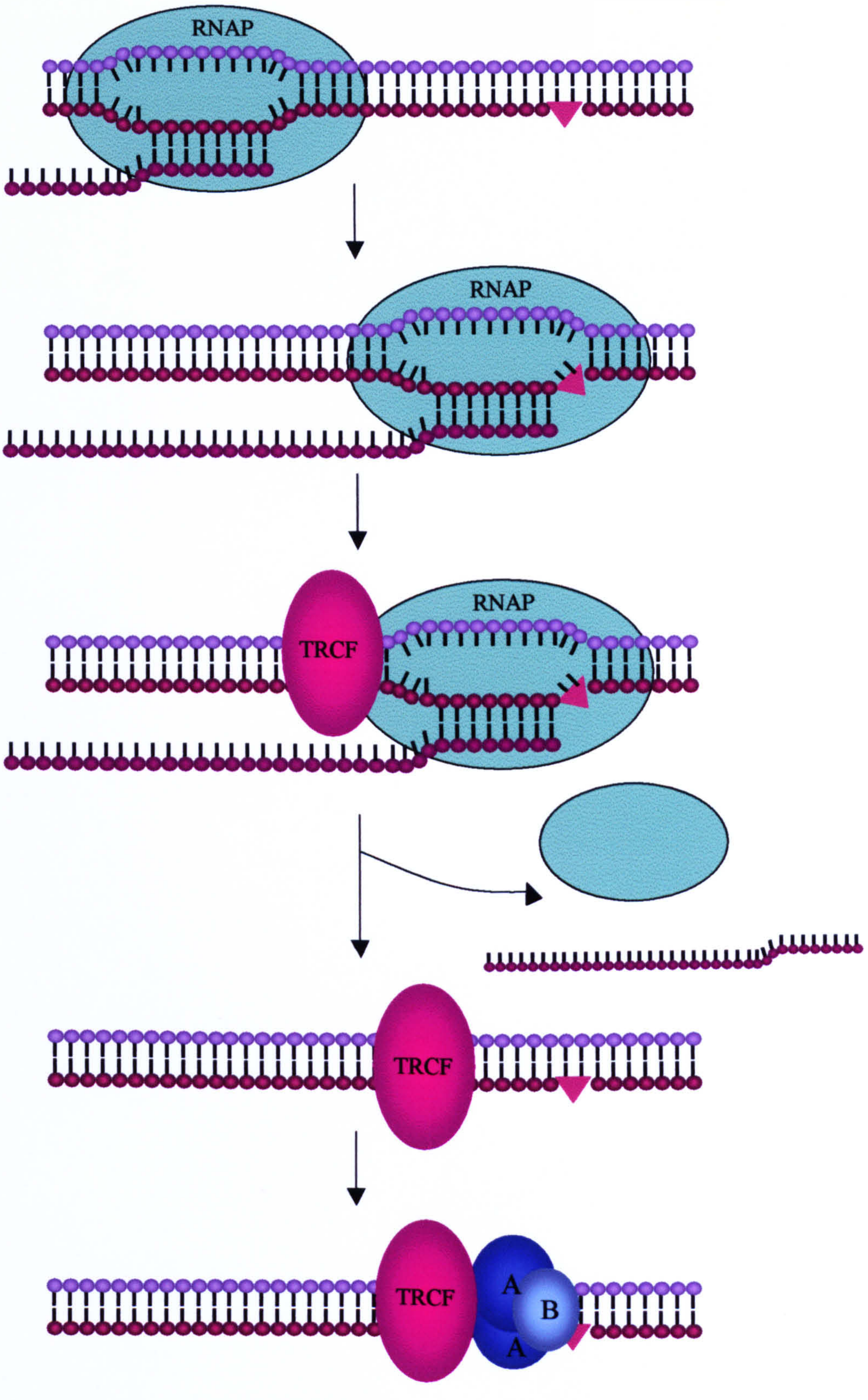


Figure 1.3. Transcription-Coupled Repair.

Transcribing RNAP becomes stalled at a thymine dimer in the template strand. A TRCF (transcription-repair coupling factor) recognises and displaces the stalled transcription complex to allow repair enzymes access to the damage. To bring about preferential repair of the transcribed strand the Uvr proteins are recruited by the TRCF.

PROPERTIES OF THE Mfd PROTEIN

The *mfd* gene encodes a protein of 1148 amino acids, present at approximately 500 copies per cell (Selby and Sancar, 1993a). The purified protein possesses ATPase, RNAP-binding, DNA-binding, UvrA-binding and RNAP displacement activities (Selby and Sancar, 1993a). A series of deletion constructs were utilised to identify domains of the Mfd protein responsible for each of these activities (figure 1.4, adapted from (Selby and Sancar, 1995a). During the course of this study, Mfd was also found to be capable of rescuing backtracked elongation complexes (Park *et al.*, 2002).

ATPase activity and helicase motifs of the Mfd protein

The amino acid sequence of Mfd contains seven helicase motifs that place it in the superfamily 2 helicase family. The ATPase activity of Mfd is modest (k_{cat}~3 min^{-1}) and is not stimulated by RNAP, single-stranded DNA or double-stranded DNA (Selby and Sancar, 1993a). Moreover, Mfd failed to exhibit strand separation activity on a variety of different substrates that were tested: DNA and RNA annealed to single-stranded phage DNA, DNA:DNA, DNA:RNA and DNA:5′-tail RNA substrates (Selby and Sancar, 1995b). A lack of strand separation activity is in fact not that unusual for superfamily 2 helicases.

By primary sequence comparison, proteins containing seven conserved 'helicase' motifs have been divided into three superfamilies and two smaller families (Gorbalenya and Koonin, 1993) (figure 1.5). The conserved motifs form a core domain and variable spacers/inserts between the motifs may determine the specific function of the protein. A number of proteins have been designated helicases based on the possession of these motifs. The majority of proteins designated as helicases

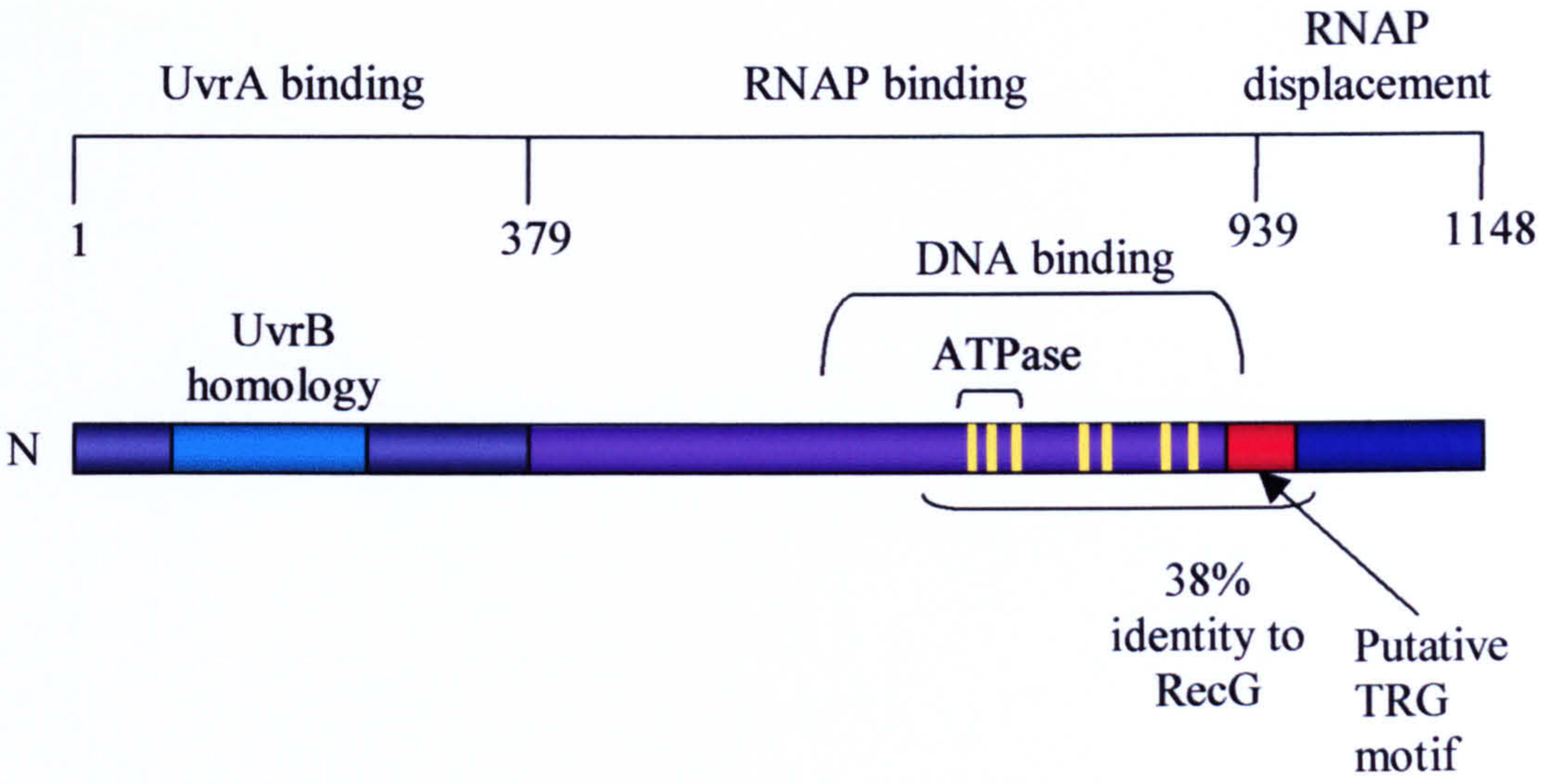


Figure 1.4. Domains of the Mfd protein.

The Mfd protein from *E. coli* is 1148 amino acids long. Amino acids 82-219 share homology with a region of UvrB while amino acids 541-983 share homology with the helicase domains of RecG. The protein contains seven superfamily 2 helicase motifs and downstream of these motifs there is a putative TRG motif. Deletion studies have mapped UvrA binding activity to amino acids 1-378, RNAP binding to 379-571 and DNA binding to 571-939. The final 210 amino acids are required for ATP-dependent RNAP displacement.

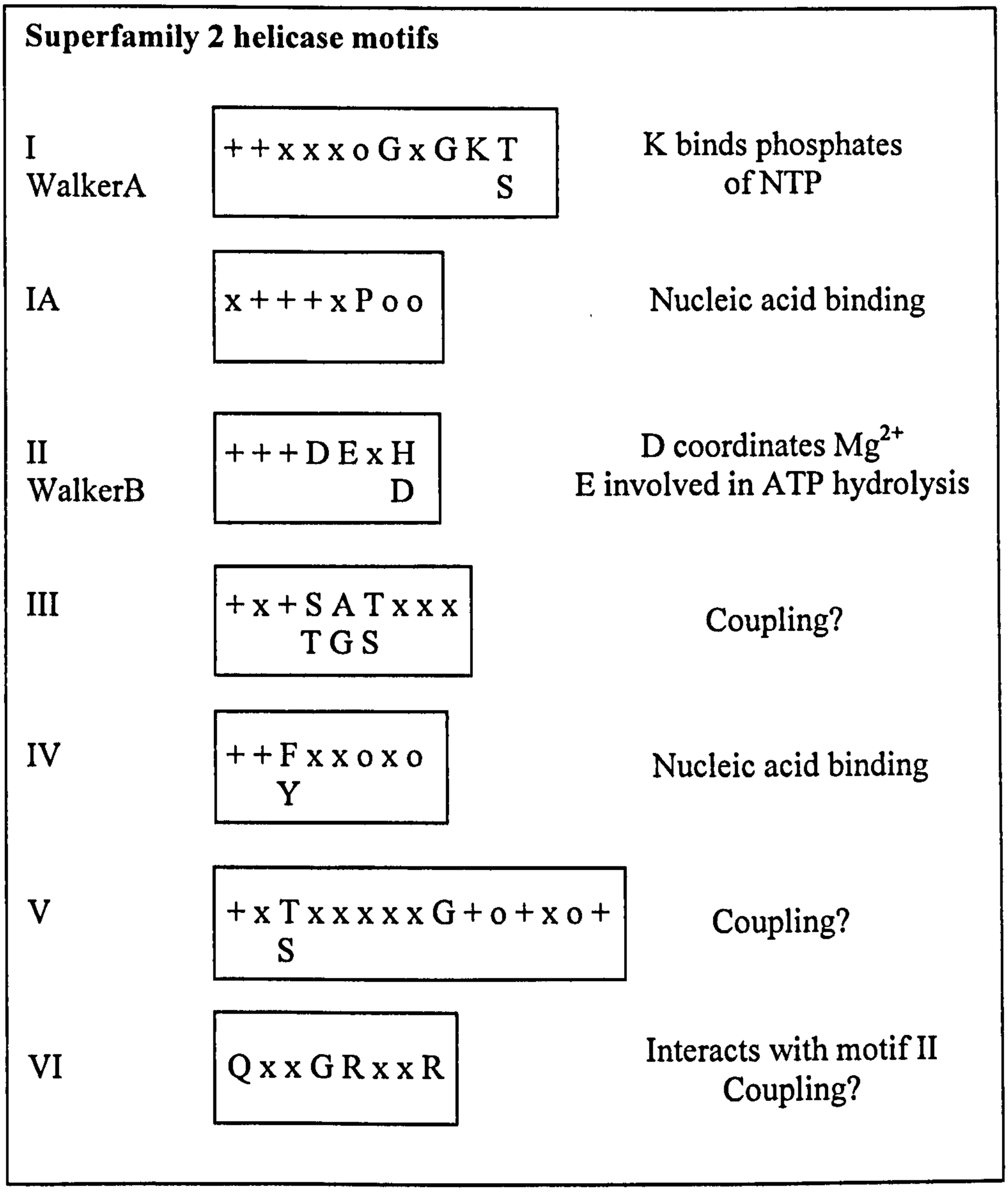


Figure 1.5. Superfamily 2 helicase motifs.

The consensus sequences for the helicase motifs of superfamily 2 helicases are shown above. x represents any amino acid, + represents a hydrophobic amino acid and o represents a hydrophilic amino acid. Motifs I and II are involved in ATP binding and hydrolysis, while the Q of motif VI interacts with H from motif II.

belong to superfamilies 1 and 2, which include RNA and DNA helicases from viruses, bacteria, archaea and eukaryotes with both 5′ to 3′ and 3′ to 5′ directionality. Within these superfamilies are a number of subfamilies consisting of proteins with significant sequence similarities. Mfd is a superfamily 2 helicase and is a member of the RecG subfamily. Other superfamily 2 subfamilies include the DEA(D/H) box RNA helicases, UvrB and the SWI/SNF2 chromatin remodelling family.

Structural information was first obtained for the SF1 helicase PcrA (Subramanya *et al.*, 1996). Structures later became available for RepA and the SF2 helicases HCV (NS3) (Yao *et al.*, 1997), eIF4A (Caruthers *et al.*, 2000), UvrB (Theis *et al.*, 1999), RecQ (Bernstein *et al.*, 2003) and RecG (Singleton *et al.*, 2001). In each of these structures the helicase motifs are located in two parallel α-β domains that possess the same fold as the RecA protein i.e. 5 parallel β strands and tandem helices. The second helicase domain possesses a RecA fold but lacks an ATP-binding motif. The nucleotide-binding pocket is located in a cleft between the two helicase domains. The most highly conserved motifs are motifs I and II, also known as the Walker A and B sequences, which are characteristic of NTPases, and are responsible for ATP binding.

Motif I forms a structure called the P-loop with the amino group of the conserved lysine binding the phosphates of the nucleotide and the hydroxyl group of the serine or threonine of motif I ligating a magnesium ion. The carboxyl group of the aspartate of the Walker B motif coordinates a magnesium ion while the glutamic acid acts as a catalytic base during ATP hydrolysis. Other residues of motif II appear to form contacts with the second helicase domain, e.g. in the UvrB and HCV crystal structure the histidine of the DEAH motif II interacts with a conserved glutamine within motif VI. The function of the other helicase motifs is not yet

understood in detail but some appear to be involved in coupling ATP binding and hydrolysis to movement on nucleic acids or in nucleic acid binding itself (reviewed in (Caruthers and McKay, 2002; Hall and Matson, 1999). Motifs Ia and IV appear to have a role in nucleic acid binding, while the functions of motifs III and V are not entirely understood but appear to be more family specific and may be important in coupling. ATP binding and hydrolysis appear to result in a change of conformation within the helicase domains that is coupled to nucleic acid binding and release.

Traditionally helicases were thought of as enzymes that moved along nucleic acids utilising the energy of ATP hydrolysis to cause separation of the two strands of a duplex. It has become clear that many proteins that possess helicase motifs do not carry strand separation activity but rather the helicase domains act as a motor for motion on DNA. Domains attached to the motor domain control the function of that particular helicase. Whilst the ATPase activity of SF1 helicases is stimulated by ssDNA, several SF2 helicases have been shown to be stimulated by dsDNA. In general it appears that while SF1 enzymes interact with nucleic acids by contacting the bases, SF2 helicases contact the backbone negating the absolute requirement for strand separation. Both the SWI/SNF2 family (Richmond and Peterson, 1996) and ISW family of chromatin remodelling enzymes (Whitehouse *et al.*, 2003); (Fitzgerald *et al.*, 2004) belong to superfamily 2. Both of these families appear to be responsible for altering the position of nucleosomes by acting as DNA-dependent translocases, suggesting a function for helicases not only in duplex destabilisation but also in destabilisation of protein-DNA interactions. In principle this could be achieved either by contacting both the DNA and the protein, or alternatively by affecting binding of the protein purely by altering the DNA conformation.

The structure of the RecG protein

Mfd is located in the same helicase subfamily as the RecG protein. Amino acids 598-968 of Mfd from *E. coli* are 38% identical to amino acids 266-645 of *E. coli* RecG, the region that encompasses the seven helicase motifs (Selby and Sancar, 1993b). The RecG protein is able to branch migrate a synthetic Holliday junction in an ATP-dependent manner (Lloyd and Sharples, 1993) and can catalyse the conversion of stalled replication forks to Holliday junctions (McGlynn and Lloyd, 2002). These junctions can be branch migrated in order to allow repair of a lesion or replication fork bypass, especially in the template for leading strand synthesis.

The structure of RecG from *Thermatoga maritima* has been solved in complex with a synthetic replication fork substrate and ADP (Singleton *et al.*, 2001) (figure 1.6 A). The protein is composed of three domains: a large N-terminal domain that binds to the DNA junction and two helicase domains. The helicase domains are N-terminally linked to the junction-binding domain through a long α helix, and a short helix at the C-terminus crosses back over this long α helix. The structure sheds light on how RecG catalyses branch migration of Holliday junctions. It is suggested that opening and closing of the cleft between the helicase domains upon ATP binding and hydrolysis is coupled to DNA translocation. It is proposed that as the helicase motor domain translocates along the double-stranded arm ahead of the junction, the DNA behind will be pulled over the junction-binding domain. The junction-binding domain contains a wedge that splits the incoming duplex DNA and stabilises the orphaned bases. The synthetic DNA junction present in the crystal structure does not contain sufficient duplex DNA ahead of the fork to determine how this it is contacted by the helicase domains. If the path of the DNA present were simply extended it

Figure 1.6. Mechanism of RecG function.

A) The crystal structure of *T. maritima* RecG in complex with ADP and a synthetic junction. The DNA is shown in green, the N-terminal junction-binding domain is shown in blue and the two helicase domains are red and magenta. The TRG motif is shown in cyan and ADP is shown in yellow. B) The model for the conversion of a collapsed replication fork to a Holliday junction by RecG. ATP binding and hydrolysis may cause a change in the orientation of the helicase domains, bringing about translocation along the double-stranded template arm. It was proposed that a swinging arm within the TRG motif contacts the DNA and acts as a lever. Dissociation of ADP causes the helicase domains to reset. Translocation on the template arm pulls the DNA over a wedge domain in the junction-binding domain, causing branch migration.

would result in a clash with the second of the two helicase domains, however a modest change in the path of the DNA would allow it to run across the surface of these domains.

Recently an additional motif has been identified in RecG, positioned downstream of the helicase motifs, named the Translocation in RecG (TRG) motif (Mahdi *et al.*, 2003). The sequence of this motif is very highly conserved between RecG proteins from different species and similar sequences also occur in the other member of the RecG helicase subfamily, Mfd. In the crystal structure of RecG the 36 amino acid motif forms a helical hairpin followed by a loop. Mutation of the TRG motif in RecG has indicated that it is involved in coupling ATP binding and hydrolysis to translocation along double-stranded DNA (Mahdi *et al.*, 2003). Using this information a model was proposed for the role of the TRG motif in coupling ATP hydrolysis to DNA translocation (figure 1.6 B). Within the helical hairpin are two conserved arginine residues that are closely juxtaposed. The presence of two positively charged arginine side chains in close proximity is energetically unfavourable, however this conformation may be stabilised by hydrogen bonds to a glutamate residue within helicase motif VI located at the N-terminal end of an α helix. An arginine residue also located within motif VI contacts the phosphate of ATP. Therefore ATP hydrolysis may cause movement in motif VI resulting in a change in the relative orientation of the two helicase domains and this change in conformation could be transmitted to the TRG hairpin. The position of the two helices could alter and in turn this movement could change the position or conformation of the following loop. The authors propose that this loop may act either as a DNA-binding swinging arm, levering the dsDNA over the surface of the protein or as a ratchet to capture DNA moved by a different mechanism.

RNAP-binding by the Mfd protein

In order to displace RNAP, Mfd must be able to recognise stalled elongation complexes. Rather than binding nucleic acid structures unique to transcription bubbles, Mfd binds most tightly to double-stranded DNA, suggesting that it may target transcription complexes by binding to RNAP rather than by binding to the nucleic acid scaffold of the transcription bubble. Although RNAP did not bind to an Mfd affinity column, pull-down assays showed that Mfd does interact with RNAP, albeit weakly (Selby and Sancar, 1995a). Mfd fused to maltose binding protein (MBP-Mfd) pulled down both core and holoenzyme, signifying that σ binding does not interfere with the binding of Mfd. Amino acids 379-571 are required for interaction with RNAP since MBP fusions to both amino acids 1-571 and 379-1148 were able to pull-down RNAP. Amino acids 379-571 share 25% identity with the N-terminal region of the CarD-like family of transcription factors (Padmanabhan *et al.*, 2001), leading to the possibility that this domain is an RNAP interaction module. The C-terminal 22% of the protein (939-1148) that by sequence analysis contains a putative leucine zipper failed to pull-down RNAP.

In a protein-protein interaction map of *Helicobacter pylori* a segment of Mfd bound to the β-β′ subunits of RNAP, which are fused as a single polypeptide in this species (Rain *et al.*, 2001). Yeast two-hybrid analysis of this segment of *E. coli* Mfd (amino acids 472-603) named the RID (**R**NAP **I**nteraction **D**omain) localised the interaction with RNAP to β, and then further to the first 142 amino acids of β (Park *et al.*, 2002). Sections of this region of β are expected to be close to where the upstream DNA exits the elongation complex.

DNA-binding by the Mfd protein

Gel retardation assays demonstrated that DNA binding by Mfd requires the binding of ATP, whilst ATP hydrolysis abolished the formation of a stable Mfd-DNA complex (Selby and Sancar, 1995a). An Mfd mutant with a substitution within the Walker A motif (Mfd KN634), that binds ATP but has dramatically reduced ATP hydrolysis activity, was able to bind DNA stably in the presence of ATP. The presence of, the non-hydrolysable ATP analog, ATPγS also stimulated formation of a stable nucleic acid-Mfd complex. It was established that Mfd binds most strongly to nucleic acid structures that contain double-stranded DNA, with a reduced binding affinity for single-stranded DNA and no binding was detected to single-stranded RNA (Selby and Sancar, 1995a). In order to displace stalled elongation complexes, Mfd requires at least 26 bp of accessible double-stranded DNA upstream of RNAP (Park *et al.*, 2002). At least 90 bp of DNA downstream of a lesion is necessary for transcription-coupled repair on linear templates (Selby and Sancar, 1995b), indicating the importance of DNA-binding by Mfd for its function.

The DNA-binding domain of Mfd was mapped to the region between amino acids 571 and 938 by the construction of a number of truncated Mfd proteins (Selby and Sancar, 1995a). Mfd that had been deleted at the N-terminus (571-1148) was still able to form stable complexes with DNA in the presence of ATPγS. Likewise a C-terminally truncated version of Mfd, termed Mfd_{trunc}, which was formed by fusion of the *lacZ* α-peptide to amino acid 938, was still able to bind DNA. However, Mfd fragments encoding amino acids 1-571 or 896-1148 did not form stable complexes with DNA.

The binding of Mfd to double-stranded DNA was examined using DNaseI footprinting on 3′ labelled fragments that varied in length but contained identical 3′

ends (Selby and Sancar, 1995a). Alternating regions of protection and hypersensitivity were observed, a pattern associated with wrapping of DNA around a protein. Protection occurred in the same positions on all of the different templates consistent with a sequence or structure within the template exerting phasing.

UvrA-binding by the Mfd protein

To fulfil its role as a transcription-repair coupling factor Mfd must be able to recruit the NER enzymes. Analysis of the primary sequence of Mfd identified a 138 amino acid region (amino acids 82-219) that shares 22% identity with amino acids 114 -251 of *E. coli* UvrB. On this basis it was predicted that this region contacts UvrA (Selby and Sancar, 1993a). Residues 1-378 of Mfd fused to MBP were able to pull-down UvrA, supporting the proposal that this region recruits UvrA (Selby and Sancar, 1995a). When different deletions of Mfd were passed over a UvrA affinity column only the 1-378 construct bound. The fact that the full-length protein did not bind could potentially be due to the UvrA interaction domain being more sterically accessible when other regions of the protein are removed. In contrast, binding occurred when UvrA was passed over an Mfd column, confirming that the two full-length proteins can interact (Selby and Sancar, 1993a). When a mixture of UvrA and UvrB was applied to this column no UvrB was retained, suggesting that UvrB and Mfd may bind to the same site on UvrA.

RNAP-displacement by the Mfd protein

Mfd-mediated displacement of elongation complexes was shown to be ATP-dependent (Selby and Sancar, 1993a). Displacement of elongation complexes by Mfd also occurred in the presence of dATP but was inhibited by the addition of ATPγS, demonstrating the requirement for ATP/dATP hydrolysis (Selby and Sancar,

1993a). Templates containing a psoralen monoadduct in the template strand caused blockage of transcription by either *E. coli* RNAP or T7 RNAP, visualised by DNaseI footprinting. Upon addition of purified Mfd, the *E. coli* RNAP was released from the adduct site. T7 polymerase was not displaced, nor were initiation complexes demonstrating that *E. coli* Mfd specifically displaces *E. coli* elongation complexes. Mfd was also able to dissociate elongation complexes that were stalled *in vitro* by a DNA-bound protein (Selby and Sancar, 1995b). The Mfd protein from *B. subtilis* was capable of displacing *E. coli* stalled elongation complexes, although less efficiently than the *E. coli* Mfd protein (Ayora *et al.*, 1996).

The length of transcripts released upon addition of Mfd corresponded to termination at the site of the block to transcription. An N-terminally truncated Mfd that contained amino acids 379-1148 was able to displace elongation complexes stalled by nucleotide starvation, however Mfd_{trunc} that lacks the C-terminal 210 amino acids was unable to dissociate RNAP (Selby and Sancar, 1995a). The region 938-1148 is therefore required for RNAP dissociation.

Native EMSAs (electrophoretic mobility shift assays) were also used to follow elongation complex displacement and provided additional information on RNAP binding (Selby and Sancar, 1995a). An Mfd protein that was unable to displace elongation complexes, $Mfd_{trunc,}$ still bound stably to an elongation complex and produced a supershifted band. The Mfd_{trunc} protein also supershifted an initiation complex in an ATP-independent manner. Despite being defective in ATP hydrolysis, and therefore unable to displace elongation complexes, an Mfd protein containing a mutation in its Walker A motif did not cause retardation of either initiation or elongation complexes. This may be due to the complexes being

insufficiently stable to withstand the relatively harsh conditions experienced during electrophoresis.

In principle there are four different mechanisms by which Mfd-mediated elongation complex displacement could occur (figure 1.7). Mfd could act on either the (1) nascent RNA (like the Rho protein) to induce termination or (2) upstream DNA. Alternatively the Mfd could (3) utilise downstream DNA or (4) induce a conformational change in the elongation complex to bring about displacement. These mechanisms are not necessarily mutually exclusive e.g. Mfd may require upstream DNA for its action but could also cause a change in the conformation of RNAP during termination. In order to terminate transcription, Mfd must be able to decrease the stability of the stalled complex either by affecting interactions in the RNA-binding site, the DNA-binding site or the hybrid-binding site or by affecting the stability of the transcription bubble.

During the course of this study, protein roadblocks were used to examine the nucleic acid requirements for elongation complex displacement (Park *et al.*, 2002). Elongation complexes were spread along a DNA template containing a recognition site for the restriction endonuclease *Eco*RI. An *Eco*RI mutant, that binds tightly to its recognition site but does not cleave DNA, was incubated with the stalled elongation complexes blocking access to different regions of upstream DNA. Only when there was at least a 26 bp gap between the downstream edge of the *Eco*RI mutant and the upstream edge of the elongation complex did Mfd-mediated transcript release occur. This implies that Mfd requires DNA immediately upstream of the stalled elongation complex for displacement. Removal of all but 3 or 4 bp of downstream DNA from the elongation complex did not affect Mfd-mediated EC displacement and demonstrated that this DNA is dispensable for RNAP dissociation

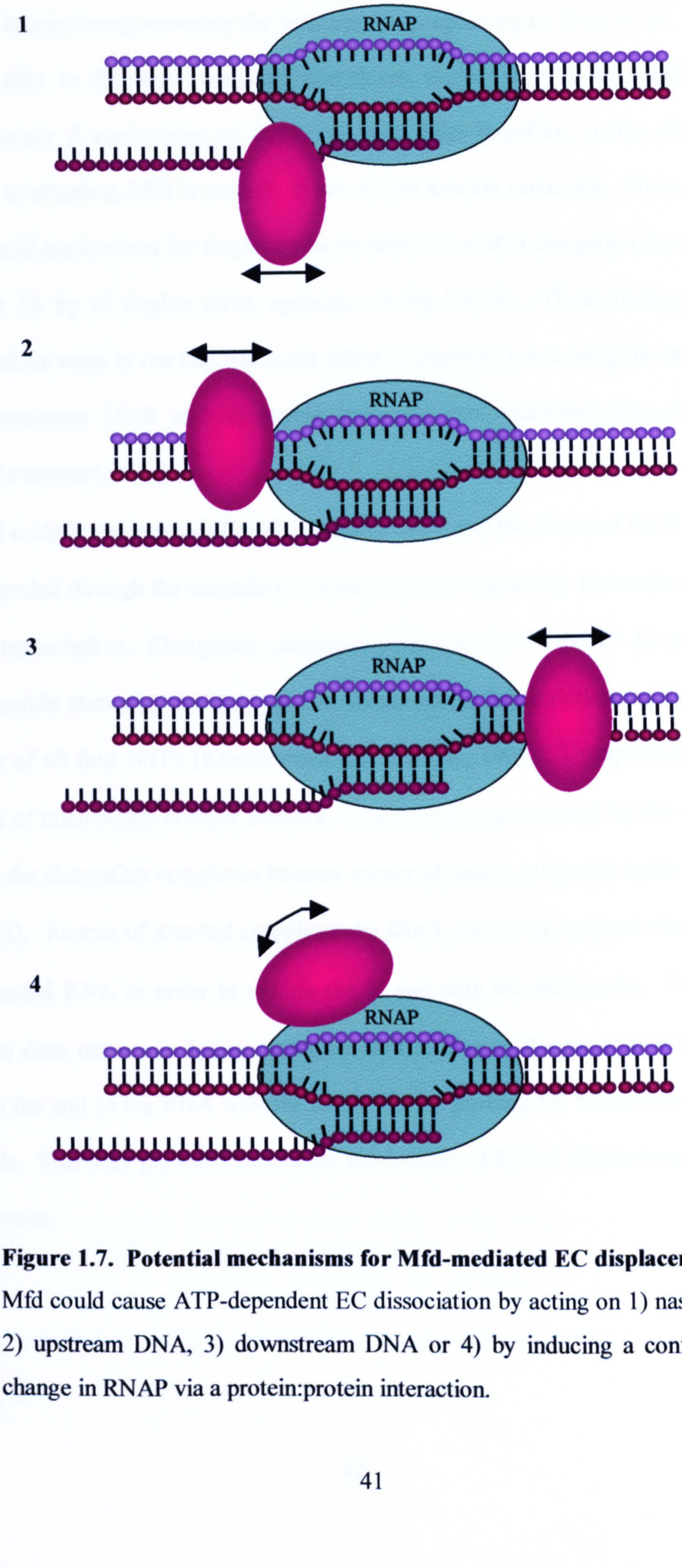


Figure 1.7. Potential mechanisms for Mfd-mediated EC displacement

Mfd could cause ATP-dependent EC dissociation by acting on 1) nascent RNA, 2) upstream DNA, 3) downstream DNA or 4) by inducing a conformational change in RNAP via a protein:protein interaction.

by Mfd, despite being necessary for transcription-coupled repair (Park *et al.*, 2002). Mfd is able to displace elongation complexes stalled at +20, in which only approximately 5 nucleotides of RNA are accessible; therefore, unlike Rho and intrinsic termination, Mfd is unlikely to act via the nascent transcript. The minimal nucleic acid requirement for displacement by Mfd of a stalled elongation complex is therefore 26 bp of duplex DNA upstream of the RNAP. These findings were confirmed by work in our lab. where the effect of physically removing the upstream and downstream DNA and accessible transcript was examined (Dr. A.Smith, personal communication).

Arrested complexes, in which RNAP has backtracked and the 3′ end of the RNA has been extruded through the secondary channel, require stimulatory factors in order to resume transcription. Elongation complexes stalled at +27 on the T7 A1 promoter by nucleotide starvation are arrested and cannot restart transcription purely by the addition of all four NTPs (Komissarova and Kashlev, 1997). Unexpectedly when addition of nucleotides to these arrested complexes is accompanied by the addition of Mfd, the elongation complexes became transcriptionally competent again (Park *et al.*, 2002). Rescue of arrested complexes by GreA and GreB involves cleavage of the extruded RNA in order to realign the 3′ end with the active site. Transcript cleavage does not occur during rescue by Mfd leading to the suggestion that Mfd realigns the end of the RNA with the active site by pushing the backtracked RNAP forwards. This may provide clues to the mechanism of RNAP displacement by the Mfd protein.

Roles of Mfd *in vivo*

Along with its role in transcription-coupled repair, the Mfd protein has also been associated with catabolite repression (Zalieckas *et al.*, 1998) and Nun-mediated transcription termination (Washburn *et al.*, 2003). Catabolite repression of the *hut* and *gnt* operons (Zalieckas *et al.*, 1998) and the *dra-nupC-pdp* (Zeng *et al.*, 2000) operon from *Bacillus subtilis* was increased by the presence of the Mfd protein. Both of these operons contain *cre* sites within the coding region that bind the CcpA repressor protein in the presence of glucose. The repression of these operons was greater in mfd^+ cells than in mfd^- cells. Therefore it seems likely that transcription complexes stalled by CcpA protein roadblocks within these operons are displaced by the Mfd protein. In *B. subtilis*, unlike in *E. coli,* the Mfd protein also appears to be involved in homologous recombination (Ayora *et al.*, 1996).

The Nun protein of bacteriophage HK022 competes with the λ protein N for binding to *nutL* and *nutR* sites of in the λ nascent transcript and prevents λ growth by terminating transcription at these sites to inhibit transcription of downstream genes (Kim and Gottesman, 2004). In an *in vitro* system Nun caused arrest of the elongation complex but it remained stably associated with the template. Mfd was shown to induce release of Nun-arrested elongation complexes both *in vitro* and *in vivo*, indicating that it is the host factor responsible for Nun-mediated termination (Washburn *et al.*, 2003).

Therefore it appears that Mfd may have a general role in termination of transcription of elongation complexes that have become stalled or backtracked, whether the stall has been induced by nucleotide starvation, the presence of a DNA lesion, a DNA-bound protein acting as a physical roadblock or by a phage protein that arrests transcription.

CSB and Rad26 proteins are involved in TCR in eukaryotes

Mutation of the CSB protein, or its yeast homolog Rad26, has been demonstrated to abolish TCR in eukaryotic cells and cause Cockayne's syndrome, which is characterised by growth retardation, neural retardation and photosensitivity. CSB is a 168 kDa protein and belongs to the SWI/SNF2 helicase family and as such possesses DNA-stimulated ATPase activity but no strand separation helicase activity. CSB binds to DNA, even in the absence of nucleotide and also contacts the transcriptional machinery by binding to TFIIH. Despite the apparent similarities with the bacterial TRCF, Mfd, CSB did not disrupt an RNA polymerase II elongation complex stalled at a pyrimidine dimer (Selby and Sancar, 1997b). It appears therefore that CSB does not function in the same manner as Mfd. As a member of the Snf2 family of chromatin remodelling proteins, CSB could remodel the interaction of RNA polymerase II with DNA to allow repair. CSB has been reported to have a general effect on transcription, enhancing the rate of elongation approximately three-fold *in vitro* (Selby and Sancar, 1997a). In addition, CSB stimulated the incorporation of one extra nucleotide at a thymine dimer. Whilst the photosensitivity of Cockayne's patients can be accounted for by a lack of TCR other symptoms may be linked to transcriptional defects.

Aims of the study

The first objective of this work was to create a readily assayed phenotype for *mfd*$^-$ cells. Prior to this study the only phenotype of *mfd*$^-$ cells was mild UV sensitivity; a means of easily selecting *mfd*$^-$ cells over *mfd*$^+$ cells was required in order to screen for substitutions in Mfd that abolish Mfd function. Development of such an *in vivo* assay for displacement of RNAP by Mfd would also allow quantification of the

effect of Mfd. In particular I intended to determine whether the putative TRG motif of Mfd is required for RNAP displacement and to examine, using site-directed mutagenesis, whether residues that are essential for a functional TRG motif in RecG are also necessary for Mfd function. Despite ATP-dependent elongation complex displacement being a well-documented activity of the Mfd protein, the molecular mechanism by which this occurs is unknown. Identification of residues involved in RNAP dissociation would help elucidate the mechanism by which Mfd carries out this function. In addition to contributing to our understanding of the mechanism of TCR, establishing a mechanism for termination of transcription by Mfd would provide insights into termination and elongation complex stability in general.

The study also aimed to identify key residues involved in the Mfd interaction with RNA polymerase, again by employing both random and site-directed mutagenesis. The intention was to purify any Mfd proteins that contained substitutions that resulted in defective Mfd function *in vivo* and to examine the properties of these proteins *in vitro*. The ability of the purified proteins to hydrolyse ATP, to bind DNA and to displace stalled elongation complexes could be tested. Finally I planned to characterise the DNA-binding properties and nucleotide hydrolysis activity of the wild-type protein.

CHAPTER 2

MATERIALS AND METHODS

STRAINS AND MEDIA

Bacterial Strains

AB1157	*thr1, ara14, leuB6, Δ(gpt-proA)62, lacY1, tsx33, qsr' ⁻, glnV44(AS), galK2(Oc), λ⁻, Rac0, hisG4(Oc), rfbD1, mgl51, rpoS396(Am), rpsL31(str^r), kdgK51, xylA5, mtl1, argE3(Oc), thi1, mfd$^+$,* F⁻ (Dewitt and Adelberg, 1962)
UNCNOMFD	Derivative of AB1157 in which the *NsiI-NsiI* fragment of the *mfd* gene has been replaced by a kanamycin encoding cassette. Therefore genotype is as AB1157 except *mfd⁻, kan^r*. (Selby and Sancar, 1993a)
XL1-Blue	*supE44, hsdR17, recA1, endA1, gyrA46, thi, relA1, lac⁻* F' [*proAB$^+$, lacI^q, lacZΔM15*Tn10 (*tet^r*)] (Stratagene)
RLG4651	M182 (*Δ(lacIOPZYA)X74, galU, galK, rpsL, Δ (ara-leu), Sm^r*) carrying λ prophage with *lacP1*::*lacZ* fusion (Savery *et al.*, 2002)

Yeast Strain

Y190	*MATa, ura3-52, his3-200, ade2-101, lys2-801, trp1-901, leu2-3, 112, gal4Δ, gal80Δ, cyh^{r}2,* LYS2::GAL1$_{UAS}$-HIS3$_{TATA}$-HIS3, MEL1 URA3::GAL1$_{UAS}$-GAL1$_{TATA}$-*lacZ* (Flick and Johnston, 1990)

Bacterial and Yeast growth media

LB (Luria-Bertani)	10 g tryptone	
	5 g yeast extract	
	10 g NaCl	per litre adjusted to pH 7
M9 minimal medium	6 g Na_2HPO_4	
	3 g KH_2PO_4	
	1 g NH_4Cl	
	0.5 g NaCl	per litre
Supplemented with	0.2% glucose	
	0.2% casaminoacids	
	0.005% $CaCl_2$	
	2 mM $MgSO_4$	
	2 µg/ml thiamine	
SOC	20 mM glucose	
	20 g tryptone	
	5 g yeast extract	
	0.5 g NaCl	per litre adjusted to pH 7
2X YT	10 g tryptone	
	10 g yeast extract	
	5 g NaCl	per litre adjusted to pH 7
YPD	10 g yeast extract	
	20 g peptone	per litre
Supplemented with	2% glucose	
	100 µg/ml ampicillin	

SD	6.7 g yeast nitrogen base without amino acids per litre	
Supplemented with	2% glucose	
	1X drop-out mix (LT⁻)	
	100 µg/ml ampicillin	

10X Drop-out mix (LT⁻)	L-Isoleucine	300 mg/l
	L-Valine	1500 mg/l
	L-Adenine hemisulphate salt	200 mg/l
	L-Arginine HCl	200 mg/l
	L-Histidine HCl monohydrate	200 mg/ml
	L-Lysine HCl	300 mg/ml
	L-Methionine	200 mg/ml
	L-Phenylalanine	500 mg/ml
	L-Threonine	2000 mg/ml
	L-Tyrosine	300 mg/ml
	L-Uracil	200 mg/ml

For all agar plates 15 g of Difco bactoagar was added per litre.

Antibiotics

Working concentrations were: -

Ampicillin	80 µg/ml
Chloramphenicol	5 µg/ml
Kanamycin	25 µg/ml
Tetracycline	20 µg/ml

Transformation of bacteria

Chemical transformation

Preparation of $CaCl_2$ competent cells

To prepare competent cells, 1 ml of an overnight culture was used to inoculate 50 ml media containing the appropriate antibiotics and competent cells were prepared essentially by the standard protocol (Sambrook and Russel, 2001). The culture was grown until mid-exponential phase (OD_{600} ~0.5) at which time the culture was transferred to a chilled falcon tube and placed on ice for 10 minutes. The cells were harvested by centrifugation at ~2,300 xg for 5 minutes at 4°C. The cell pellet was resuspended in 20 ml of ice-cold 0.1 M $CaCl_2$ and was incubated on ice for 20 minutes. The cells were again harvested by centrifugation at ~2,300 xg for 5 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in 3 ml ice-cold 0.1 M $CaCl_2$ and 2 ml ice-cold 50% glycerol. The prepared competent cells were incubated on ice for at least 30 minutes before they were used, or were snap frozen in aliquots and stored at -80°C.

Transformation of $CaCl_2$ competent bacteria

For transformations, 10 ng of plasmid DNA or 15 µl of ligation was placed with 100 µl of competent cells on ice for 30 minutes. A heat shock was applied by placing the cells at 42°C for 2 minutes. Following the heat shock, 500 µl of LB was added immediately and the cells were allowed to recover for one hour at 37°C. The cells were pelleted by centrifugation at 16,100 xg for 30 seconds, the media was removed and the pellet was resuspended in 100 µl LB before being spread onto an agar plate containing the appropriate antibiotics.

Electroporation

Preparation of electrocompetent cells

Electrocompetent cells were prepared by the method supplied by Biorad. To make electrocompetent cells, 50 ml of media containing the appropriate antibiotics was inoculated with 1 ml of an overnight culture. The culture was grown to an OD_{600} of ~0.5 before the culture was transferred to a pre-chilled falcon tube and placed on ice for 20 minutes. The cells were harvested by centrifugation at ~2,300 xg for 15 minutes at 4°C. The cells were washed three times in 50 ml, 25 ml and then 2 ml ice-cold 10% glycerol, before the cell pellet was resuspended in 200 µl of ice-cold 10% glycerol. Cells were incubated on ice for at least 30 minutes prior to use.

Transformation by Electroporation

For electroporation, 2 µl of ligation or the desired amount of DNA was added to 40 µl of electrocompetent cells. The cells were then placed in a 0.1 cm electroporation cuvette (Biorad). Approximately 1500 V was applied to the cuvette over approximately 5 ms and 1 ml SOC was added immediately to permit recovery. The cells were incubated at 37°C for 30 minutes before being pelleted at 16,100 xg for 30 seconds. Cells were resuspended in 100 µl broth before being spread onto agar plates.

METHODS FOR DNA MANIPULATION

Unless stated otherwise chemicals were purchased from Sigma and enzymes were purchased from Roche or NEB.

Ligations, phosphatase treatment of vectors and fragments to be end-labelled and standard restriction enzyme digests were performed by standard methods (Sambrook and Russel, 2001).

Partial digests

On some occasions, incomplete digestion by restriction enzymes was necessary. For partial digests, 1 μg DNA was incubated at 37°C with 0.5 or 1 unit of restriction enzyme in a 50 μl reaction. After 0, 2.5, 5, 10, 15, 30, 60, 90, 120 and 180 minutes, 4 μl samples were taken and were stopped by the addition of 0.05 M EDTA. Samples were analysed by gel electrophoresis in a 1% agarose gel to identify the incubation time at which the maximum amount of linear product was obtained. Multiple digests were then performed under the conditions identified and the products were combined.

Gel electrophoresis

Agarose and acrylamide gels used for resolution of DNA

Plasmids and fragments larger than 1 kb were analysed on 1% agarose/TAE (40 mM Tris acetate, 1 mM EDTA) gels containing 0.5 μg/ml ethidium bromide. Prior to loading, 6X loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in 1X TBE (90 mM Tris-borate, 2 mM EDTA)) was added to samples. Fragments were sized by comparison to 250 ng of 1 kb gene ruler (MBI Fermentas). Gels were run at 110V for 2 hours before viewing on a Kodak Image Station 440 imager or UV transilluminator.

Fragments smaller than 1 kb were analysed on either 2% agarose/TAE gels containing 0.5 μg/ml ethidium bromide or on 7.5% acrylamide (Severn Biotech,

30% w/v, 37.5:1 bis acrylamide)/TBE gels. Acrylamide gels were set by the addition of 100 µl 10% APS and 10 µl TEMED per 10 ml. Agarose gels were run as above, while acrylamide gels were run at 100 V for one hour before staining for 20 minutes in 100 ml of 0.5 µg/ml ethidium bromide prior to viewing.

Denaturing gels

Sequencing reactions and DNaseI footprinting reactions were resolved on 6% polyacrylamide (Severn Biotech, 19:1 bis acrylamide), 6.8 M urea, 1X TBE gels, 0.4 mm thick. Gels were set by the addition of 100 µl 10% APS and 10 µl TEMED per 10 ml. Gels were pre-run to a constant temperature of 50°C (120W initially) before the samples were loaded. The gels were fixed in 10% methanol, 10% acetic acid before they were dried on a Biorad gel dryer for 1 hour.

SDS-PAGE gels

Proteins were resolved on SDS-PAGE gels (Sambrook and Russel, 2001) composed of a 15% acrylamide (Severn Biotech, 30% w/v 37.5:1 bis acrylamide), 0.29 M Tris pH 8.3, 0.1% SDS resolving gel and a 6% acrylamide (Severn Biotech, 30% w/v 37.5:1 bis acrylamide), 0.13 M Tris pH 6.8, 0.1% SDS stacking gel. Gels were set by the addition of 100 µl 10% APS and 10 µl of TEMED per 10 ml. Samples were mixed with 2X SDS loading buffer (100 mM TrisCl pH 6.8, 0.2% bromophenol blue, 4% SDS, 200 mM DTT) and were boiled for 2 minutes prior to loading. Samples were resolved on the gel for 1 hour at 200 V in SDS running buffer (25 mM Tris, 263 mM glycine, 1% SDS).

Gels were stained for 30 minutes in Coomassie Blue (0.2% Coomassie Blue, 50% methanol, 10% glacial acetic acid), before destaining for 15 minutes in fast destain (40% methanol, 10% acetic acid), followed by a further 15 minutes in slow destain

(10% methanol, 10% acetic acid) and changes of destain until the background staining had been removed.

Analysis of radioactive gels

Radioactive gels were dried and visualised using a phosphor image cassette (Amersham Pharmacia Biotech) and a Typhoon phosphorimager (Amersham Pharmacia Biotech). Images were analysed using ImageQuant software (Molecular Dynamics) and all gel images shown are on a linear scale.

Purification of DNA from agarose and acrylamide gels

DNA was extracted from agarose gel slices using either a GeneCleanIII kit (Q-Biogene) or a Qiagen Gel Extraction kit according to the manufacturer's protocols. Following GeneClean gel extraction, the DNA was eluted from the glass milk in 20 µl TE pH 8.0 (10 mM Tris, 1 mM EDTA). DNA was eluted from Qiagen columns in 30 µl TE.

DNA was extracted from acrylamide gel slices by electroelution. Dialysis tubing (12,000-14,000 Da MWCO, Medicell International Ltd.) was prepared by boiling for 10 minutes in 1-2 litres of 2% sodium bicarbonate, 1 mM EDTA pH 8.0. The tubing was rinsed and boiled for a further 10 minutes in 1 mM EDTA pH 8.0. The excised gel slice was placed inside the dialysis tubing along with approximately 200 µl 0.1X TBE. The dialysis bags were then placed in a mini gel tank containing 0.1X TBE and the DNA was electroeluted for 40 minutes at 30 mA. The buffer was transferred from the dialysis bag to a microfuge tube and the bag was washed with 100 µl TE, which was combined with the buffer. The DNA was concentrated by phenol chloroform extraction and ethanol precipitation before resuspension in 20 µl TE.

Phenol chloroform extraction and ethanol precipitation

A volume of phenol: chloroform: isoamyl alcohol (25: 24: 1) equal to the sample was added and was mixed vigorously for 30 seconds. The aqueous and organic phases were separated by centrifugation at 16,100 xg for 5 minutes. The upper aqueous phase was transferred to a new tube and one-tenth volume of 3 M NaAc (sodium acetate) pH 5.2 was added. Two volumes of ice-cold 100% ethanol were added and the tube was placed at -20°C for 45 minutes. The DNA was pelleted by centrifugation at 4°C, 16,100 xgfor 15 minutes and the ethanol removed. The pellet was washed with 1 ml 70% ethanol and was then centrifuged again at 4°C, 16,100 xg for a further 15 minutes. The ethanol was removed and the pellet was allowed to air dry for 10 minutes prior to resuspension in the desired volume of TE.

Kinase treatment of linear DNA fragments

DNA fragments were end-labelled at the 5′ terminus by incubation with T4 polynucleotide kinase (Roche) and 10 μCi [γ^{32}P]-ATP (NEN) for 30 minutes at 37°C in the buffer supplied by the manufacturer. Unincorporated nucleotides were removed using Bio-spin columns (Biorad) according to the manufacturer's protocol.

Preparation of plasmid DNA

By alkaline lysis method

Small-scale plasmid DNA preparation was carried out essentially by the standard procedure (Sambrook and Russel, 2001). LB (5 ml) containing the appropriate antibiotics was inoculated with a single bacterial colony. Overnight cultures were grown at 37°C with vigorous shaking. Cells from 1.5 ml of overnight culture were harvested by centrifugation for 1 minute at 16,100 xg The supernatant was removed

and the pellet was resuspended in 100 µl of ice-cold solution I (50 mM glucose, 10 mM EDTA, 25 mM Tris Cl pH 8.0) by vortexing. Following incubation in solution I for 5 minutes at room temperature, 200 µl of freshly prepared solution II (0.2 M NaOH, 1% SDS) was added and the contents mixed by inverting the tubes. Cells were lysed on ice for 5 minutes before the addition of 150 µl of solution III (3 M potassium acetate pH 4.8). The contents were mixed by inversion and were incubated on ice for 5 minutes. The precipitate was pelleted by centrifugation for 5 minutes at 16,100 xg, the supernatant was transferred to a fresh microfuge tube and 300 µl of phenol/chloroform was added. The tubes were vortexed for ~15 seconds before centrifugation for 2 minutes at 16,100 xg, 300 µl of the upper aqueous phase was transferred to a fresh tube and 600 µl of ethanol at room temperature was added. The tubes were immediately subject to centrifugation at 16,100 xg for 5 minutes to pellet the plasmid DNA. The supernatant was discarded and 1 ml of 70% ethanol was added to wash the pellet. The tubes were centrifuged again at 16,100 xg for 5 minutes before the ethanol was removed and the pellet was briefly dried in a vacuum desiccator. The DNA pellet was resuspended in 50 µl TE.

By Qiagen spin prep

DNA for sequencing was prepared using either Qiaquick Spin Miniprep or Midiprep kits according to the manufacturer's protocol.

Caesium chloride large-scale plasmid DNA preparation

For large-scale plasmid preparation, 5 ml of LB was inoculated with a single colony from an agar plate and was grown overnight. This overnight culture was used to inoculate 300 ml of LB containing the appropriate antibiotics. The culture was grown at 37°C until an OD_{600} of approximately 1–1.2 was reached. At this time

chloramphenicol was added to a final concentration of 170 μg/ml and the culture was incubated overnight at 37°C. The culture was transferred to a chilled centrifuge pot and was incubated on ice for 30 minutes. The cells were harvested by centrifugation at 4,000 xg at 4°C for 15 minutes. The pellet was resuspended in 5 ml of ice-cold solution I (50 mM glucose, 10 mM EDTA pH 8.0, 25 mM TrisCl pH 8.0). 0.5 ml of 10 mg/ml lysozyme was added, the tubes were mixed by inversion and were then incubated on ice for 5 minutes. Following this incubation, 10 ml of solution II (0.2 M NaOH, 1% SDS) was added and the tube contents were mixed by inversion. The tube was incubated on ice for a further 5 minutes before 7.5 ml of ice-cold solution III (3 M potassium acetate pH 4.8) was added. The tube contents were mixed by inversion before incubation on ice for a further 5 minutes. The precipitate that formed was removed by centrifugation at 17,600 xg at 4°C for 15 minutes, and 13 ml of isopropanol was added to the supernatant. The tube was incubated at room temperature for 15 minutes, and then the DNA was pelleted by centrifugation at 17,600 xg at 20°C for 15 minutes. The pellet was rinsed with 5 ml 70% ethanol and was recentrifuged at 17,600 xg at 20°C for 10 minutes and the supernatant was discarded. The pellet was air-dried and was resuspended in 8 ml TE to which 8.2 g of caesium chloride and then 200 μl of 10 mg/ml ethidium bromide was added. The solution was centrifuged at 5,000 xg at 20°C for 10 minutes to pellet the RNA. The supernatant was removed from below a layer of protein "scum", and was transferred to a 13.5 ml ultracentrifuge tube. Paraffin was added to fill the tube before it was sealed and was centrifuged at 181,000 xg at 15°C for at least 16 hours. The band containing supercoiled DNA was removed using a needle and was extracted 3-6 times with an equal volume of water-saturated butanol to remove the ethidium bromide. The DNA solution was dialysed against two changes of 500 ml TE. The

dialysed product was extracted with an equal volume of phenol/chloroform and the phenol/chloroform was then back-extracted with an equal volume of TE that was combined with the aqueous phase from the first extraction. The DNA was precipitated by the addition of 0.1 volumes of 3 M sodium acetate pH 5.2 and 2.5 volumes of ice-cold 100% ethanol, and incubated at -20°C for 1 hour. The DNA was harvested by 15 minutes centrifugation at 17,600 xg at 4°C. The pellet was rinsed with 10 ml of 70% ethanol and was centrifuged again at 17,600 xg at 4°C for 15 minutes. The DNA pellet was resuspended in 500 µl TE buffer pH 8.0.

Polymerase Chain Reaction

Sequences of oligonucleotides are contained in the Appendix.

Standard fragment amplification by PCR

PCR reactions contained 1 µM forward and reverse primers, 200 µM dNTPs, 10 ng template, 1X polymerase buffer and 0.5 µl Pfu DNA polymerase (Promega) in a 50 µl reaction volume. The standard protocol used began with a denaturation step at 94°C for 5 minutes after which the DNA polymerase was added. This step was followed by 30 amplification cycles of melting, annealing and elongation, typically 94°C for 1 minute, 56°C for 1 minute and 72°C for 1 minute. A final 5 minute extension was carried out at 72°C before the reactions were chilled to 4°C.

Site-directed mutagenesis PCR

Fusion PCR was used to introduce site-directed mutations (Karreman, 1998). The PCR was carried out in two stages; the first generating two PCR products, each with the desired mutation incorporated into one of the primers and the second generating

a longer product with the mutation incorporated into the middle of the PCR product. To generate the two half-length PCR products, PCR reactions were carried out using standard PCR conditions with one mutagenic primer and one wild-type primer. The PCR products were purified using a Qiagen PCR purification kit according to the manufacturer's protocol and were eluted in 30 µl water.

The two PCR products that were generated in the first step were used as template in a second PCR reaction to generate the full-length product. For the second step, 3 µl of both PCR products from the first round were used in a reaction with 1 µM of the forward and reverse wild-type primers, 200 µM dNTPs, 1X polymerase buffer and 0.5 µl Pfu or Vent DNA polymerase. The PCR protocol used was as above. The final PCR product was purified using the Qiagen PCR purification kit according to manufacturer's protocol and eluted in 30 µl TE.

Colony screen PCR

Colony screen PCR was used to screen a large number of colonies for the presence or absence of an insert or the introduction of a specific mutation, directly from bacterial colonies on an agar plate (Sambrook and Russel, 2001). A master mix containing 1X polymerase buffer, 200 µM dNTPs and 1 µM each primer was prepared and dispensed into 24.5 µl aliquots. A sterile pipette tip was touched to a colony and used to streak onto a second agar plate before the bacteria were transferred to the PCR mix by pipetting up and down. The tubes were incubated for 10 minutes at 100°C to break open the cells and release the DNA. 0.5 µl Taq was added to each tube before 30 cycles of 94°C for one minute, 50°C for 2 minutes and 72°C for 2 minutes.

Error-prone PCR

Two different error-prone PCR protocols were used to introduce random mutations. In protocol 1, 40 cycles of PCR were carried out utilising the natural error frequency of Taq (8×10^{-6} errors/bp (Cline *et al.*, 1996)). The PCR reactions used 2 μM of the forward and reverse primers, 10 ng template, 200 μM dNTPs and 0.5 μl Taq (Promega) in a 50 μl reaction in the reaction buffer supplied with the polymerase.

Error-prone PCR protocol 2 utilised Mn^{2+}, unequal dNTP concentrations and Taq in order to increase the error frequency (Greene J, 2002). The error-prone PCR reactions contained 2 μM forward and reverse primers, 10 ng template, 10 mM Tris pH 8.0, 50 mM KCl, 7 mM $MgCl_2$, 1 mM dCTP, 1 mM dTTP, 0.2 mM dATP, 0.2 mM dGTP and 0.5 μl Taq in a 50 μl reaction. Between 5 and 10 cycles of error-prone PCR were carried out before the reactions were purified with a Qiagen PCR purification kit and amplified for a further 25 cycles using high-fidelity PCR with Pfu under standard conditions.

DNA sequencing

Constructs were either confirmed by automated sequencing at MWG Biotech or the University of Dundee sequencing services or by manual sequencing as described below. Plasmid DNA prepared using QIAprep spin plasmid kit (Qiagen) was eluted from the column in 100 μl TE. 3 μl fresh 2 M NaOH was mixed with 30 μl of the double-stranded DNA and was incubated at room temperature for 15 minutes. Following this incubation, 3 μl 5 M ammonium acetate pH 4.8 was added and the tube contents were mixed briefly. Immediately 75 μl of cold ethanol was added and the samples were placed at -20°C for 15 minutes. Single-stranded DNA was pelleted by centrifugation for 15 minutes at 16,100 xg at 4°C and the pellet was washed with

1 ml 70% ethanol before being centrifuged for a further 5 minutes at 16,100 xg at 4°C. The supernatant was removed and the pellet was dried using a vacuum desiccator. The single-stranded DNA pellet was resuspended in 10 µl water and was mixed with 2 µl primer (4 pmols/µl) and 2 µl of annealing buffer from a Pharmacia T7 sequencing kit. Sequencing was carried out according to the manufacturer's protocol. The dATP labelling mix was used with 1 µl [α-^{32}P]dATP (10 µCi) (Amersham). Sequencing reactions were heated to 90°C for two minutes before 2 µl of each reaction was resolved on a 6% acrylamide denaturing gel.

PLASMIDS

Summary of plasmids

Table 2.1. Plasmid summary table (alphabetical order).

Plasmid	Description	Source
pACT-RpoB	Amino acids 1-142 of the β-subunit fused to the Gal4 activation domain of pACT. Complements the leucine auxotrophy of yeast strain Y190	Dr. A. Smith
pAR1707	Plasmid containing the T7A1 promoter from which a transcript is produced that does not require UTP until +21 if ApU is used as an initiating dinucleotide	Levin *et al.*, 1987

pAS2.1	Generates a fusion with a Gal4 DNA-binding domain when a protein of interest is cloned into the MCS. Carries *trp1* gene, *bla* gene and constitutive *adh1* yeast promoter and terminator.	Clontech
pAS-Mfd and derivatives	Amino acids 472-603 of Mfd fused to Gal4 DNA-binding domain of pAS-2.1. Confers ability to grow in tryptophan free media.	Dr. A. Smith
pBluescript KS+	Amp^r	Stratagene
pBSMfd-TRG	Used as a template for random mutagenesis PCR. Amino acids 850-997 cloned into pBluescript KS+ with *Hin*dIII, *Bam*HI ends	Alyce Merry Project student
pBSMFD-RID	RNAP interaction domain of Mfd cloned into the *Xho*I-*Eco*RV sites of pBluescript KS+	This work
pCM7	Contains the *cat* gene	Pharmacia
pCSB1	*cat* gene from pCM7 cloned into *pSRgalPcon6,* flanked either side by an optimal *lac* operator	Claire Bell (Chambers *et al.*, 2003)
pET21a	Contains $lacI^q$, Amp^r Empty vector control for pETMfd	NEB
pET21term	Intermediate in the construction of pETMfd2. λoop terminator from pSR cloned into pET21a backbone	This work
pETMfd and derivatives containing substitutions	*mfd* gene and promoter from pMFD19 cloned into pET21a backbone	This work
pETMfd2	*mfd* gene (including terminator) inserted into pET21term in opposite orientation to pETMfd	This work

pETMfd2AgeI and derivatives containing substitutions	Derivative of pETMfd2. Silent mutation within codon 481 to introduce an *Age*I site within *mfd*.	This work
pETMfd2Trunc$_{1\text{-}939}$	Stop codon introduced after codon 939 and downstream amino acids deleted in pETMfd2. Truncated within TRG motif.	This work
pETMfd2Trunc$_{1\text{-}997}$	Stop codon introduced after codon 997 and downstream amino acids deleted in pETMfd2. Truncated following short helix after TRG motif.	This work
pGL2	Contains the *luc* gene	Promega
pKA4	Optimal *lac* operator site plus a wild-type *lac* operator site 70.5 bp downstream of the *galPcon6* promoter	Kate Adams (summer student)
pKA4luc	Intermediate in construction of pRCBKA4. Promoter, Shine-Dalgarno and optimal *lac* operator site plus a wild-type *lac* operator site 70.5 bp downstream (from pKA4) inserted into pSRluc.	This work
pMFD19	*Sph*I-*Ssp*I fragment containing *mfd* gene and promoter cloned into the *Hinc*II-*Sph*I sites of pIBI25, pUC19-based	(Selby and Sancar, 1993a)
pMFD19trunc	Derivative of pMFD19. Stop codon introduced in *mfd* downstream of codon 939. Truncated within TRG motif.	This work
pRCB	Derivative of pRW50 with the *lacZYA* genes removed	Claire Bell (Chambers *et al.*, 2003)
pRCB-CAT	Roadblock reporter plasmid containing optimal *lac* operator sites either side of the *cat* gene, downstream of the *galPcon6* promoter	Claire Bell (Chambers *et al.*, 2003)

pRCBKA4	Region of pKA4luc containing promoter, operators, *luc* gene and terminator cloned into pRCB	This work
pRCBlacO$_{ID}$luc	Promoter, Shine-Dalgarno and optimal *lac* operator site from pSRgalPcon6CAT lacO#1 inserted into pRCB	This work
pRW50	Broad host range plasmid containing *lacZ* reporter, pRK2 *ori*, Tet^r	(Lodge *et al.*, 1992)
pSRc33	Plasmid containing the lacUV5 promoter from which a transcript is produced that does not require CTP until +34	Dr. A. Smith
pSRca19	Derivative of pSRc33 that will transcribe to +19 using only GTP and UTP and the initiating dinucleotide ApU	Dr. A. Smith
pSRgalPcon6CATlacO#1	Optimal *lac* operator site placed downstream of *galPcon6* promoter	Kate Adams
pSRlacO$_{ID}$luc	Promoter, Shine-Dalgarno and optimal *lac* operator site from pSRgalPcon6 lacO#1 inserted into pSRluc	This work
pSRlacUV5 (-140/63)	Contains the *lacUV5* promoter and the λoop terminator	(Savery *et al.*, 2002)
pSRluc	Intermediate in construction of pRCBKA4 and pRCBlacO$_{ID}$luc. *Luc* gene from pGL2 was silently mutated to destroy an internal *Eco*RI site before cloning into pSRgalPcon6 as *Eco*RI-*Hin*dIII fragment	This work
pUC19	Amp^r	(Yanisch-Perron *et al.*, 1985)
pUCMfd	Template for random mutagenesis PCR. Amino acids 895-997 of Mfd in pUC19 with a *Bsm*AI site at the downstream end	This work

Plasmid construction

pET series containing full-length Mfd

pETMfd

The *Eco*RI-*Sph*I fragment from pMFD19 containing the *mfd* gene and promoter (Stanley and Savery, 2003) was cloned into the *Eco*RI-*Sph*I sites of pET21a (figure 2.1). The construction of derivatives of pETMfd containing substitutions in the TRG region of Mfd is shown in table 2.2 and construction of derivatives of pETMfd containing substitutions in the Q-tip of Mfd is shown in table 2.3.

pETMfd2

A *Sty*I site was introduced downstream of the λoop terminator from pSR by PCR using primers D3421 and pSRds2 and pSRlacUV5(-140/63) as a template. The PCR product was cut with *Hin*dIII and *Sty*I and was cloned into a *Hin*dIII-*Sty*I pET21a vector to generate the intermediate plasmid pET21term.

The *mfd* gene and promoter was cloned as a *Bam*HI-*Hin*dIII fragment from pMFD19 into the *Bgl*II-*Hin*dIII sites of pET21term to generate pETMfd2. The *mfd* gene is in the opposite orientation relative to pETMfd and is terminated by the λoop terminator. The construction of plasmids containing truncations of Mfd, based on pETMfd2 is shown in table 2.3.

pETMfd2AgeI

This plasmid contains an *Age*I site introduced by silent mutation of codon 481 to allow screening of the potential RNAP contact region of Mfd in two sections of ~300 bp. The *Age*I site was introduced using fusion PCR. For the first round of PCR to

generate two partial products, MFD033 and MFD034 were used as primers for the upstream section while MFD027 and MFD032 were used for the downstream section. Both partial PCR reactions used pETMfd as the template. The primers used to generate the full-length product were MFD027 and MFD034. The PCR product was digested with *XhoI* and *PmeI* and cloned into the *Xho*I and *Pme*I sites of pETMfd2 (figure 2.1). The construction of derivatives of pETMfd2AgeI containing substitutions in the Mfd RID is shown in table 2.4.

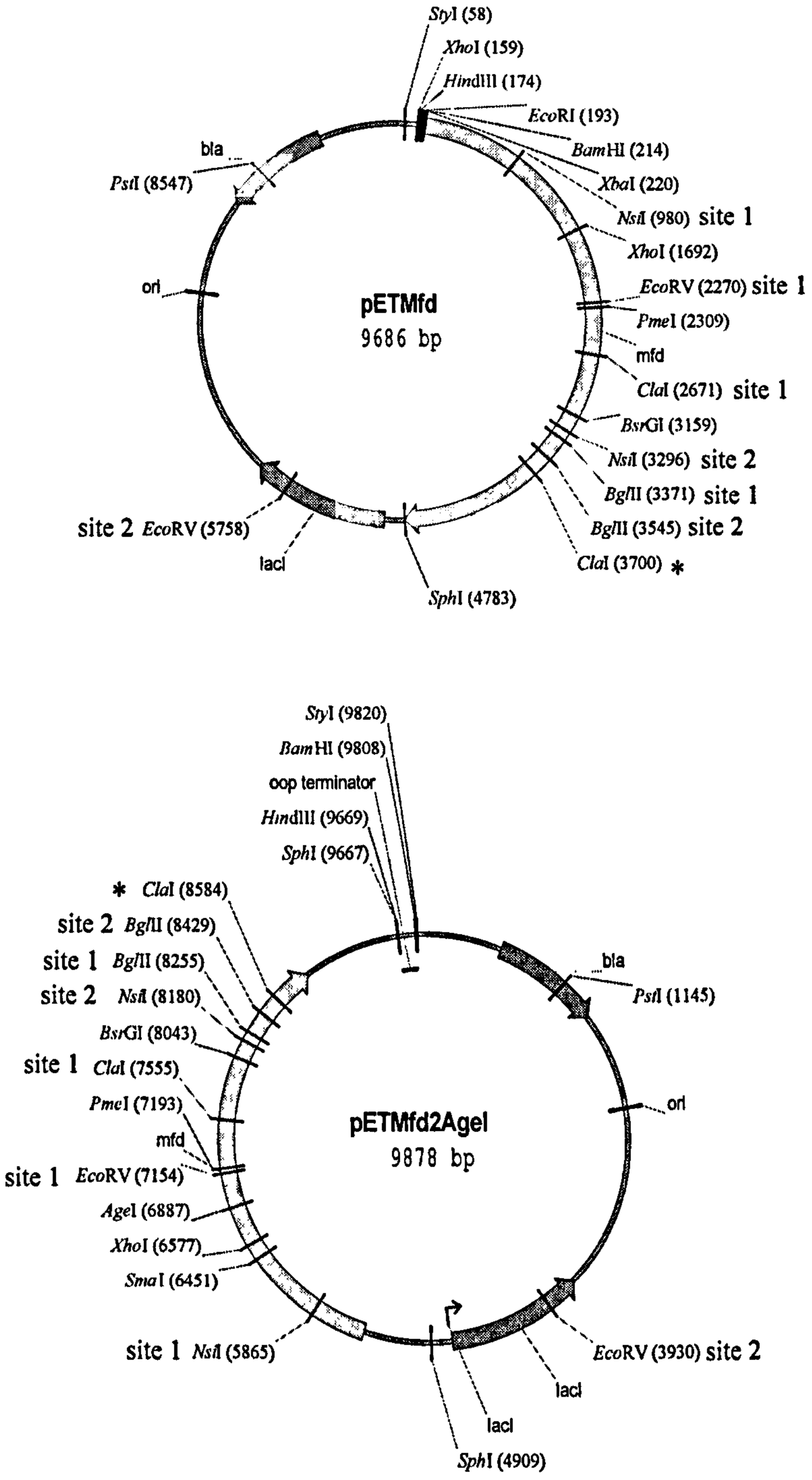


Figure 2.1. Restriction maps of pETMfd and pETMfd2AgeI.

Restriction sites used during construction of derivatives of pETMfd and pETMfdAgeI, and for restriction mapping during random mutagenesis screening are shown. The *Cla*I site marked by an asterisk is blocked by dam methylation, making the other *Cla*I site effectively unique.

Plasmid name	Mutagenesis method	Primer pairs	Template for PCR	Restriction fragment	Vector
pETMfdQL898	Random (2)	(Bluescript 1, Bluescript 2)	pBS-MFDTRG	*Bsr*GI-*Bam*HI	*Bsr*GI-*Bgl*II pETMfd
pETMfdRA929	Site-directed	(M13 1233, **MFD013**) (**MFD012**, M13 D9979)	pUC-Mfd	*Nsi*I-*Bsm*AI	*Nsi*I (site 2)-*Bgl*II (site 2) pETMfd
pETMfdHL948	Random (1)	(M13 D9979, M13 1233)	pUC-Mfd	*Nsi*I-*Bsm*AI	*Nsi*I (site 2)-*Bgl*II (site 2) pETMfd
pETMfdRA953	Site-directed	(MFD005, **MFD003**) (**MFD006**, MFD002)	pMFD19	*Bgl*II	*Bgl*II pMFD19 then *Eco*RI-*Sph*I fragment was subcloned into pET21a
pETMfdQA963	Site-directed	(MFD005, **MFD025**) (**MFD024**, MFD002)	pETMfd	*Bgl*II	*Bgl*II pETMfd
pETMfdGD956	Random (2)	(Bluescript 1, Bluescript 2)	pBS-MFDTRG	*Bsr*GI-*Bam*HI	*Bsr*GI-*Bgl*II pETMfd
pETMfdGV960	Random (2)	(Bluescript 1, Bluescript 2)	pBS-MFDTRG	*Bsr*GI-*Bam*HI	*Bsr*GI-*Bgl*II pETMfd

Table 2.2. Plasmids encoding substitutions in the TRG region of Mfd.

The method of mutagenesis, the primers used, the template used for PCR, the restriction fragment containing the mutation and the vector the fragment was cloned into are shown. Primers are shown as pairs and mutagenic primers are shown in bold. For substitutions generated by random mutagenesis, it is shown whether error-prone PCR protocol 1 or 2 was used. pETMfdQL898 was isolated by Alyce Merry, an undergraduate student under my supervision.

Plasmid name	Mutagenesis method	Primer pairs	Template for PCR	Restriction fragment	Vector
pMFD19Trunc	Site-directed	(MFD001, **MFD005**)	pMFD19	*Nsi*I-*Hin*dIII	*Nsi*I (site 2)-*Hin*dIII pMFD19
pETMfd2Trunc$_{1\text{-}939}$	-	-	-	*Xho*I- *Hin*dIII	*Xho*I-*Hin*dIII pETMfd2
pETMfd2Trunc$_{1\text{-}997}$	-	-	-	*Bgl*II	*Bgl*II pETMfd2Trunc$_{1\text{-}939}$
pETMfdDA604	Site-directed	(MFD030, **MFD029**) (**MFD028**, MFD031)	pETMfd	*Eco*RV-*Cla*I	*Eco*RV (site 1)-*Cla*I (site 2) pETMfd
pETMfdQA605	Site-directed	(MFD030, **MFD027**) (**MFD035**, MFD031)	pETMfd	*Eco*RV-*Cla*I	*Eco*RV (site 1)-*Cla*I (site 2) pETMfd

Table 2.3. Plasmids encoding truncations of Mfd or substitutions in the Q-tip of Mfd.

The method of mutagenesis, the primers used, the template used for PCR, the restriction fragment containing the mutation and the vector the fragment was cloned into are shown. Primers are shown as pairs and mutagenic primers are shown in bold. For substitutions generated by random mutagenesis, the error-prone PCR protocol that was used is indicated.

Plasmid	Method of mutagenesis	Primers	Restriction fragment
pETMfd2RA472	Site-directed	(MFD034, **MFD040**) (**MFD039**, MFD029)	*Xho*I-*PmeI*
pETMfd2EA476	Site-directed	(MFD034, **MFD042**) (**MFD041**, MFD029)	*Xho*I-*PmeI*
pETMfd2HA485	Site-directed	(MFD034, **MFD044**) (**MFD043**, MFD029)	*Xho*I-*PmeI*
pETMfd2HA488	Site-directed	(MFD034, **MFD046**) (**MFD045**, MFD029)	*Xho*I-*PmeI*
pETMfd2RA492	Site-directed	(MFD034, **MFD048**) (**MFD047**, MFD029)	*Xho*I-*PmeI*
pETMfd2LR499	Random (2)	(MFD034, MFD029)	*Age*I-*Pme*I
pETMfd2EA500	Site-directed	(Bluescript2, **MFD050**) (**MFD049**, Bluescript1)	*Xho*I-*Eco*RV
pETMfd2EA507	Site-directed	(MFD034, **MFD052**) (**MFD051**, MFD029)	*Xho*I-*PmeI*
pETMfd2KA518	Site-directed	(MFD034, **MFD054**) (**MFD053**, MFD029)	*Xho*I-*PmeI*
pETMfd2HL527	Random (2)	(MFD034, MFD029)	*Age*I-*Pme*I
pETMfd2RA531	Site-directed	(MFD034, **MFD056**) (**MFD055**, MFD029)	*Xho*I-*PmeI*
pETMfd2KA544	Site-directed	(MFD034, **MFD058**) (**MFD057**, MFD029)	*Xho*I-*PmeI*
pETMfd2RA552	Site-directed	(MFD034, **MFD060**) (**MFD059**, MFD029)	*Xho*I-*PmeI*
pETMfd2RA554	Site-directed	(MFD034, **MFD062**) (**MFD061**, MFD029)	*Xho*I-*PmeI*
pETMfd2KA556	Site-directed	(MFD034, **MFD064**) (**MFD063**, MFD029)	*Xho*I-*PmeI*
pETMfd2DA563	Site-directed	(MFD034, **MFD066**) (**MFD065**, MFD029)	*Xho*I-*PmeI*
pETMfd2EA567	Site-directed	(MFD034, **MFD068**) (**MFD067**, MFD029)	*Xho*I-*PmeI*

Table 2.4. Plasmids encoding substitutions in the RID of Mfd.

pETMfd2AgeI was used as the PCR template and as the vector, with the exception of pETMfd2EA500, for which pBSMfd-RID was used as the PCR template and *Xho*I-*Eco*RV (site 1) pETMfd2 was used as the vector. Primers are shown in pairs, with primers encoding the desired substitution shown in bold. The method of mutagenesis is indicated including whether error-prone PCR protocol 1 or 2 was used.

Templates for random mutagenesis

pUCMfd

The region of the *mfd* gene encoding amino acids 895-997 was cloned into pUC18 in order to permit mutagenesis of this region by error-prone PCR using M13 primers that anneal within pUC18. PCR was used to introduce a *Bsm*AI site adjacent to the *Bgl*II site2. *Bsm*AI cuts 1 and 4 bp away from its recognition sequence and therefore its recognition site was positioned such that when a digest with *Bsm*AI was performed, a *Bgl*II compatible end was generated. The primers MFD005 and MFD011 were used to generate a PCR product with a *Bsm*AI site positioned in this manner at the 3′ end. The PCR product was cloned in a blunt end ligation into the *Sma*I site of pUC18.

pBSMfd-TRG

The region of the *mfd* gene encoding amino acids 850-997 was cloned into pBluescript KS+ in order to use it as a template for error-prone PCR. PCR was used to generate an insert with *Hin*dIII-*Bam*HI ends that was cloned into the polylinker of pBluescript KS+.

pBSMFD-RID

The *Xho*I-*Eco*RV fragment of pETMfd2AgeI, which contains the RNAP interaction domain of Mfd, was cloned into the polylinker of pBluescript KS+ to generate pBSMFD-RID.

Reporter constructs

pSRluc

pSRluc was an intermediate in the construction of pRCBKA4 and pRCBlacO$_{ID}$luc. First, silent site-directed mutagenesis was carried out on the firefly luciferase gene of pGL2 (Promega) to remove the internal *Eco*RI site. Primers were used that introduced an *Eco*RI site followed by an *Nco*I site at the 5′ end of the luciferase gene product and a *Hin*dIII site at the 3′ end. To generate the upstream product, luc1 and luc3 were used as primers while the downstream product was produced using luc4 and luc2 as primers. Both reactions used pGL2 as the template. The full-length product was generated using luc1 and luc2 as primers. This PCR product was cloned into pSRgalPcon6 as an *Eco*RI-*Hin*dIII fragment, resulting in the plasmid pSRluc.

pRCBlacO$_{ID}$luc

PCR was carried out on pSRgalPcon6CATlacO#1 to amplify the promoter region and ideal *lac* operator plus the Shine-Dalgarno sequence and also to introduce an *Nco*I site at the ATG start codon. D5431 and pSRSDNcoI were used as primers. The PCR product was cloned into pSRluc as an *Eco*RI-*Nco*I fragment to generate pSRlacO$_{ID}$luc. The *Eco*RI-*Bam*HI fragment containing the promoter, operator, *luc* gene and terminator of pSRlacO$_{ID}$luc was then cloned into pRCB to generate pRCBlacO$_{ID}$luc.

pRCBKA4

PCR was carried out on pKA4 to amplify the promoter region, ideal *lac* operator and the wild-type operator 70.5 bp downstream (centre-to-centre), plus the Shine-Dalgarno sequence. In addition an *Nco*I site was introduced at the ATG start codon.

D5431 and pSRSDNcoI were used as primers. The PCR product was cloned into pSRluc as an *Eco*RI-*Nco*I fragment, to generate pSRKA4luc. The *Eco*RI-*Bam*HI fragment containing the promoter, operators, *luc* gene and terminator of pSRKA4luc was cloned into pRCB to generate pRCBKA4.

Yeast two-hybrid plasmids

pAS-LR499Mfd, pAS-HA485Mfd and pAS-RA554Mfd express the Gal4 DNA-binding domain fused to amino acids 472-603 of Mfd. PCR primers containing an *Nco*I site (5′MfdRID) and a *Bam*HI site (3′MfdRID) were used to amplify the appropriate region of Mfd using pETMfd2LR499, pETMfd2HA485 or pETMfd2RA554 as a template. The PCR products were digested using *Nco*I and *Bam*HI and were cloned into the multiple cloning site of pAS-2.1, generating pAS-LR499Mfd, pAS-HA485Mfd and pAS-RA554Mfd.

IN VIVO REPORTER ASSAYS

Assay of β-galactosidase from *E. coli* colonies

As a secondary screen during random mutagenesis, β-galactosidase assays (Miller, 1972) were performed to identify colonies containing functional Lac repressor. Z-buffer (10 mM KCl, 1 mM $MgSO_4$, 60 mM Na_2HPO_4, 30 mM NaH_2PO_4, 50 mM β-mercaptoethanol) was supplemented with 0.003% SDS and 0.66 mg/ml ONPG. The Z-buffer/SDS/ONPG mix was dispensed into 1 ml aliquots to which 30 µl chloroform was added. A sterile toothpick was used to streak a colony onto an LB plate containing ampicillin, for subsequent retrieval of DNA from positive clones, before being placed into the β-galactosidase reaction mix. The tube was vortexed for

10 seconds and the appearance of a bright yellow colour was observed for colonies that lacked *lacI*^{q} after approximately 2 minutes at room temperature.

Growth of cultures for *cat* and *luc* roadblock repression assays

Competent UNCNOMFD (*mfd*) cells containing the roadblock reporter construct pRCB-CAT were prepared. In the case of experiments examining transdominant phenotypes, AB1157 (*mfd*$^+$) cells were used. The cells were transformed with the appropriate pET-based Mfd plasmid or the empty pET21a vector and were plated onto M9 agar plates containing kanamycin, tetracycline and ampicillin. An overnight culture from a single colony was grown in M9 media containing kanamycin, tetracycline and ampicillin. The overnight culture was diluted to an OD_{600} of 0.05 in M9 media containing kanamycin, tetracycline and ampicillin and was grown to an OD_{600} of approximately 0.5 before the cells were harvested by centrifugation of 1.5 ml culture at 16,100 xg for 5 minutes. Where indicated, 0.5 mM IPTG was added to the culture during subculture.

Preparation of cell extracts for CAT assays

The cell pellet was resuspended in 150 µl of Bugbuster (Novagen) and was incubated for 15 minutes at room temperature. The suspension was then centrifuged 16,100 xg for 20 minutes at 4°C. The supernatant was split into two 75 µl aliquots and the extracts were snap frozen in liquid nitrogen before they were stored at -80°C.

Determination of protein concentration

The total protein concentration of the cell extracts was determined by Bradford assay (Bradford, 1976). Samples were prepared in a volume of 800 µl to which 200 µl of Biorad protein assay reagent was added. Samples were incubated for at least 5

minutes at room temperature before their absorbance was read at 595 nm against a blank of 800 µl water and 200 µl protein assay reagent. A calibration curve was prepared using 1.25, 2.5, 3.75 and 5 µg BSA.

Quantification of CAT activity using Quant-T-CAT assay kit

The Quant-T-CAT assay kit (Amersham Pharmacia Biotech) allows quantification of CAT activity by measuring the transfer of tritiated acetyl groups from acetyl-coA to chloramphenicol. The chloramphenicol used is biotinylated to allow binding to streptavidin-coated beads for removal of free acetyl-coA. Appropriate dilutions of cell extracts were prepared in Bugbuster and 20 µl of the dilution was transferred to a reaction tube. Extracts from cells containing functional Mfd were diluted approximately 2-fold whilst extracts from cells containing no Mfd or defective Mfd were diluted approximately 15-fold. The extracts from induced cultures i.e. those grown in the presence of IPTG were diluted approximately 60-fold. A CAT standard curve was also prepared containing 0, 10, 20, 30, 40, 50 and 60 mU of CAT in Bugbuster in a total volume of 20 µl.

The diluted sample (20 µl) was added to 5 µl of a reaction master mix resulting in a final reaction volume of 25 µl that contained 2 µM biotinylated chloramphenicol, 120 mM Tris pH 7.8 and 0.25 µCi [^{3}H] acetyl coenzyme A. The reaction was allowed to proceed for 45 minutes at 37°C before being stopped by the addition of 500 µl of the supplied wash buffer containing 1 mg of streptavidin-coated beads. The reaction was placed at room temperature for 5 minutes in order to allow binding to the beads. The beads and bound chloramphenicol were pelleted by centrifugation at 16,100 xg for 5 minutes before 490 µl of the supernatant was discarded and the pellet was washed with 500 µl of wash buffer to remove unincorporated tritiated

acetyl CoA. The tube was centrifuged for 5 minutes at 16,100 xg and another 490 µl of supernatant was removed. The pellet was washed again with 500 µl wash buffer and was centrifuged for a final 5 minutes at 16,100 xg. 500 µl of supernatant was removed and the pellet was resuspended in 1 ml Elgastat (ultra-high purity) water and was transferred to a scintillation vial. 10 ml of scintillation fluid (Perkin Elmer) was added and the activity of each sample was measured by counting for 10 minutes in a scintillation counter (Packard). The counts from the tube containing Bugbuster alone were subtracted as background. Each assay was carried out on three different colonies.

Preparation of cell extracts for luciferase assays

Cultures were grown to an OD_{600} of ~0.5 under the conditions indicated above, at which point two 0.5 ml aliquots were harvested by centrifugation at 16,100 xg for 2 minutes. The cell pellets were resuspended in 90 µl M9 media containing 10 µl 1 M K_2HPO_4, pH 7.8 and 20 mM EDTA. Cell suspensions were snap frozen in liquid nitrogen and allowed to thaw in a room temperature water bath. The lysed cell suspension was mixed with 300 µl of lysis mix (1X CCLR (25 mM Tris-phosphate pH 7.8, 2 mM DTT, 2 mM 1,2-diaminocyclohexane-N, N, N′, N′-tetraacetic acid, 10% glycerol, 1% Triton X-100), 2.5 mg/ml BSA, 1.25 mg/ml lysozyme) and incubated for 10 minutes at room temperature. Cell debris was pelleted by centrifugation at 16,100 xg for 1 minute and the supernatant was transferred to a fresh tube. To assay the luciferase activity of the samples, 20 µl lysate was added to 100 µl luciferase assay reagent (Promega) and readings were taken after 3 minutes incubation at room temperature using a luminometer (Berthold Technologies).

YEAST TWO-HYBRID ASSAYS

Preparation of competent yeast

A modified version of the Clontech protocol for LiAc preparation of competent yeast was followed. Several Y190 yeast colonies were resuspended in 1 ml of YPD media. The 1 ml suspension was used to inoculate 10 ml YPD, which was then incubated overnight at 30°C at 250 r.p.m.. The overnight culture was diluted to an OD_{600} of 0.2 in 60 ml YPD and was incubated for approximately 3 hours at 30°C, in orbital shakers at 250 r.p.m., until the yeast reached exponential phase (OD_{600} 0.4-0.6). Cells were harvested by centrifugation at 180 xg for 5 minutes. Each pellet was resuspended in 40 ml water and was centrifuged at 180 xg for 5 minutes. The pellet was then resuspended in 4 ml TE/LiAc (1X TE, 100 mM LiAc) and centrifuged for a further 5 minutes at 180 xg. Each pellet was resuspended in 300 µl TE/LiAc and competent yeast were transformed within an hour of preparation.

Yeast transformation

Yeast were transformed with 0.5 µg pACT-RpoB and 0.5 µg of either pAS-Mfd, pAS-HA485Mfd, pAS-LR499Mfd or pAS-RA554Mfd. 5 µl 10 mg/ml herring sperm DNA was added to the DNA to be transformed; the herring sperm DNA acted as a carrier and had been boiled for 20 minutes and cooled on ice prior to use. The DNA was mixed with 50 µl competent yeast and 300 µl PEG/LiAc (1X TE, 100 mM LiAc, 40% PEG) was added. Tubes were incubated at 30°C for 30 minutes without shaking. Following this incubation, 35 µl DMSO was added and the contents of the tube were mixed by swirling. The yeast were incubated at 42°C for 15 minutes to apply a heat shock before being placed on ice for 2 minutes. The yeast were pelleted by centrifugation at 180 xg for 5 minutes and the pellet was resuspended in 300 µl

1X TE. The transformation was spread onto SD LT⁻ selective agar and was incubated at 30°C for 3-5 days to allow growth of yeast colonies.

Liquid β-galactosidase assays on yeast cell extract

To assay the β-galactosidase activity (Miller, 1972) of transformed yeast, 5 ml SD LT⁻ media was inoculated with a single yeast colony and was grown overnight at 30°C, in an orbital shaker at 250 r.p.m.. The OD_{600} of the overnight culture was measured. Cells from 3 ml of overnight culture were harvested and the pellet was resuspended in 400 µl TT buffer (50 mM Tris pH 7.5, 0.5% Triton X-100). The suspension was frozen in liquid nitrogen for 1 minute and was allowed to thaw at room temperature before the debris was pelleted by centrifugation at 16,100 xg for 1 minute. Assay reaction mixtures contained 200 µl of the yeast protein extract and 600 µl Z-buffer; 200 µl TT buffer was used in place of cell lysate as a blank reaction. Reactions were started by the addition of 200 µl ONPG (4 mg/ml) and were incubated at 37°C until some of the tubes began to turn yellow. The reactions were stopped by the addition of 400 µl 1 M Na_2CO_3. The tubes were centrifuged at 16,100 xg for 1 minute and the OD_{420} measured. β-galactosidase activity was calculated as OD_{420}/OD_{600} of overnight culture and was expressed relative to the activity of the wild-type construct. Assays were performed on three different colonies.

PURIFICATION OF MFD PROTEINS

Mfd proteins were purified from UNCNOMFD cells containing the appropriate Mfd plasmid expressing *mfd* from its own promoter. Cells were grown in 1 l 2X YT in the presence of 25 µg/ml kanamycin and 80 µg/ml ampicillin for approximately 8

hours at 37°C. The cells were harvested by centrifugation at 4,200 xg at 4°C for 20 minutes. The pellet was resuspended in 20 ml of lysis buffer (50 mM Tris pH 7.5, 120 mM NaCl, 5% glycerol (v/v), 10 mM EDTA and 5 mM DTT) and following treatment with 0.3 mg/ml lysozyme for 15 minutes on ice, 10 µl of 10 mg/ml RNase and 10 µl 1 U/ml DNase were added and the cell suspension was incubated for a further 20 minutes on ice. The cells were lysed by two 15 second bursts of sonication at 40% amplitude using a Sonics Vibracell sonicator. Debris was pelleted by centrifugation at 7,800 xg at 4°C for 20 minutes. The supernatant was passed through a 0.45 µm filter before being loaded onto a 5 ml HiTrap Blue Sepharose column (Amersham). The protein was eluted using a 0.05-1 M KCl gradient in TGED buffer (10 mM Tris pH 7.5, 1 mM EDTA, 5% glycerol (v/v) and 1 mM DTT) over 5 column volumes, at a flow rate of 1 ml/min. The fractions that contained Mfd were diluted 10-fold in TGED or dialysed overnight into TGED, before being loaded onto a 5 ml heparin column (Amersham). A 0.05-1 M KCl gradient in TGED buffer over 5 column volumes was used to elute the protein at a flow rate of 1 ml/min and the fractions containing Mfd were diluted 10-fold in TGED before being loaded onto a 1 ml MonoQ column (Amersham). The protein was eluted with a 0.05-1 M KCl gradient in TGED buffer over 50 column volumes at a flow rate of 1 ml/min and the purified protein was dialysed against Mfd storage buffer (10 mM Tris pH 8.0, 50% glycerol (v/v), 1 mM EDTA, 2 mM DTT and 200 mM KCl) in a 10,000 MWCO Slide-a-lyser cassette (Pierce). The protein concentration was determined using the Bradford assay and was stored at -20°C. During purification the purity of the Mfd protein was followed by running fractions on SDS-PAGE gels.

ATPase ASSAYS

ATPase activities were determined using an assay that monitored the release of P_i from [γ-^{32}P]-ATP (modified from (Tombline and Fishel, 2002). Unhydrolysed ATP was removed by adsorption onto activated charcoal. ATP hydrolysis assays were carried out in a reaction volume of 10 µl containing 2 mM ATP and 1 µCi [γ-^{32}P] ATP in repair buffer (40 mM HEPES pH 8.0, 100 mM KCl, 8 mM $MgCl_2$, 4% glycerol, 5 mM DTT, 100 µg/ml BSA). Reactions were started by the addition of 1 µM Mfd and were incubated at 37°C in a PCR machine with a heated lid to prevent evaporation. The reactions were stopped at different time points by the addition of 400 µl ice-cold 30 mg/ml charcoal in 40 mM HCl. The tubes were incubated on ice for 3 hours to allow adsorption of unhydrolysed ATP to the charcoal. The tubes were centrifuged at 16,100 xg for 30 minutes and two 50 µl samples of the supernatant were taken from each tube and added to 4 ml water for Cerenkov counting. The rate of background hydrolysis was established from reactions without Mfd. In order to calculate the specific activity the total counts were measured by omission of charcoal from the acid used to stop the reactions.

DNA-BINDING ASSAYS

DNA-binding assays were either carried out on the 250 bp *Eco*RI-*Bam*HI fragment from pSRc33 or a 32 bp fragment formed by annealing the complementary oligonucleotides MFD012 and MFD013. The 250 bp fragment was labelled at each 5′ end by kinase treatment and MFD012 was also labelled at the 5′ end. The labelled oligo was heated to 95°C for 5 minutes with equimolar amounts of unlabelled

MFD013 before cooling to room temperature over several hours to generate a 32 bp double-stranded oligonucleotide.

Unless indicated otherwise, assays analysing binding to the 250 bp fragment contained 100 nM purified Mfd protein and 0.4 nM labelled fragment. Where indicated ATP or ATPγS were also present at a final concentration of 2 mM. Reactions analysing binding to 32 bp dsDNA contained 0.4 nM double-stranded oligo, 0, 50, 100, 200, 300 or 500 nM Mfd and 2 mM ATPγS and were incubated at 37°C for 25 minutes.

Reactions were run, without the addition of loading buffer, on 8 mM magnesium acetate, 5% acrylamide, 1X TAE gels at 150 V for approximately 3 hours at 4°C.

ELONGATION COMPLEX DISPLACEMENT ASSAYS

Displacement of elongation complexes by Mfd was analysed using EMSAs. Elongation complexes were stalled on three different templates by nucleotide starvation and contained either an end-labelled DNA template or a labelled transcript due to incorporation of labelled nucleotides during transcription.

Reactions using labelled DNA template

Using pAR1707 fragment

Stalled elongation complexes were formed by nucleotide starvation on a fragment that requires UTP incorporation at +2 and +21; therefore elongation complexes could be stalled at +20 using the dinucleotide ApU to initiate transcription and by omission of UTP from the transcription mix. A 529 bp *Rsa*I-*Sma*I fragment from pAR1707 was end-labelled and initiation complexes were formed by incubating 0.4 nM labelled fragment with 5 nM RNAP holoenzyme (Epicentre) in repair buffer for 15

minutes at 37°C. Where indicated, 80 μM ApU, 0.8 mM ATP, 8μM CTP and 8μM GTP were added and the reactions were incubated at 37°C for 15 minutes to form elongation complexes. To prevent further transcription initiation 5.5 μg/ml rifampicin was added prior to Mfd being added at a final concentration of 400 nM. The reactions were incubated at 37°C for 15 minutes before being run on a 4.5% acrylamide /1X TBE gel at 200V at 4°C.

For experiments to determine the time course of Mfd-mediated displacement, 0.4 nM fragment was incubated with 5 nM RNAP (Epicentre) in repair buffer for 15 minutes at 37°C. Unstable complexes were removed by the addition of 10 μg/ml heparin before addition of ApU, ATP, GTP and CTP at final concentration of 80 μM, 1.7 mM, 8 μM and 8 μM respectively. The reaction was incubated for 15 minutes at 37°C. Mfd was added at a final concentration of 250 nM and aliquots were taken and loaded onto gels under tension.

Using pSRc33 fragment

Elongation complexes stalled by nucleotide starvation were formed on the 252 bp *Eco*RI-*Bam*HI fragment from pSRc33. CTP is not required until +34 in this transcript therefore exclusion of CTP from the transcription reaction results in elongation complexes halted at +33. The fragment was end-labelled using T4 polynucleotide kinase and [γ^{32}P]-ATP. 10 nM RNAP holoenzyme (a gift from Dr. A. Smith, His-tagged at the C-terminal end of the β-subunit and purified according to (Niu *et al.*, 1996)) was incubated in repair buffer with 0.4 nM labelled fragment for 15 minutes at 37°C to form initiation complexes. To remove any RNAP that had not formed an open complex 10 μg/ml heparin was added. Transcription was initiated by the addition of 2 mM ATP, 10 μM GTP and 10 μM UTP and reactions

were incubated for 15 minutes at 37°C to allow stalled elongation complexes to form. Mfd was then added at the concentration indicated and the reactions were incubated at 37°C.

For time course experiments a sample was taken before the addition of Mfd and then samples were taken at the indicated time points and were loaded onto a 4.5% acrylamide/1X TBE gel under tension (70V). The gels were run at 4°C at 200V for approximately 4 hours.

For Mfd titration experiments the reactions were incubated for 5 mins following the addition of Mfd prior to running the gels under the above conditions.

Using pSRca19 fragment

The *Eco*RI-*Bam*HI fragment from pSRca19 can be transcribed until +19 in the presence of ApU, GTP and UTP. Therefore transcription elongation complexes that are stalled due to a requirement for CTP at +20 can be formed in the absence of ATP and CTP. The fragment was prepared by the same method as the fragment used for DNaseI footprinting, so that end-labelling using T4 polynucleotide kinase and [γ^{32}P]-ATP caused only the non-transcribed strand to be radiolabelled. To form initiation complexes 0.4 nM labelled fragment was incubated with 10 nM RNA polymerase (His-tagged at the C-terminus of the β-subunit) for 15 minutes at 37°C. To form elongation complexes, 10 μg/ml heparin, 100 μM ApU, 10 μM GTP and 10 μM UTP were added to the reactions and they were incubated at 37°C for 15 minutes to form stalled elongation complexes. Where indicated 250 nM Mfd was added and the reactions were started by the addition of 2 mM nucleotide. The reactions were incubated for 15 minutes at 37°C before being loaded onto a 4.5% acrylamide/1X TBE gel under tension (70V). The gels were run at 60V at 4°C overnight.

Reactions using labelled RNA transcript

Using pAR1707 fragment

Elongation complexes stalled by nucleotide starvation were radiolabelled by incorporation of labelled CTP into the nascent transcript. Initiation complexes were formed by incubating 10 nM RNAP holoenzyme (Epicentre) with 2 nM *Rsa*I-*Sma*I fragment from pAR1707 in repair buffer at 37°C for 15 minutes. First rifampicin and then ApU, ATP and GTP were added at final concentrations of 20 μg/ml, 100 μM, 2 mM and 10 μM respectively. Unlabelled CTP was added to a final concentration of 2 μM along with 5 μCi [α^{32}P]CTP. The reactions were incubated at 37°C for 15 minutes to form elongation complexes before Mfd was added to a final concentration of 400 nM. The reactions were incubated for 15 minutes at 37°C before being run on a 4.5% acrylamide/1X TBE gel. To prevent unincorporated labelled nucleotide entering the gel the cathode buffer was exchanged once the samples had entered the gel, and the gel was run at 200V at 4°C for approximately 4 hours.

Using pSRca19 fragment

Elongation complexes stalled at +20 by CTP starvation were labelled by the incorporation of radioactive UTP into the transcript. To form initiation complexes, 2 nM *Eco*RI-*Bam*HI fragment of pSRca19 was incubated in repair buffer with 10 nM RNAP (His-tagged at the C-terminus of the β-subunit) for 15 minutes at 37°C. To form elongation complexes, 10 μg/ml heparin, 100 μM ApU, 10 μM GTP, 2 μM UTP and 5 μCi [α^{32}P] UTP were added and the reactions were incubated for 15 minutes at 37°C to form stalled elongation complexes. Where indicated 250 nM Mfd and 2 mM nucleotide was added. The reactions were incubated for 15 minutes

at 37°C before being loaded onto a 4.5% acrylamide/1X TBE gel under tension. To prevent unincorporated labelled nucleotide entering the gel the cathode buffer was exchanged once the samples had entered the gel, and the gel was run at 200V at 4°C for approximately 4 hours.

DNaseI FOOTPRINTING

To prepare a footprinting fragment 100 μg of pSRc33 caesium chloride maxiprep was cut with either *Eco*RI or *Bam*HI, depending which end of the fragment was to be labelled, and was treated with phosphatase. The DNA was purified by extraction with phenol/chloroform and ethanol precipitation and was then digested with the other enzyme. The fragment was end-labelled by kinase treatment before 4 nM labelled fragment was incubated in a reaction volume of 20 μl with 2 mM ATPγS and 500 nM wild-type Mfd or mutant Mfd. Reactions were placed at 37°C for 25 minutes before 3 μl of 1 μg/ml DNaseI (Pharmacia) was added. The reaction was stirred for 15 seconds and then placed at room temperature for a further 30 seconds before being stopped by the addition of 200 μl DNaseI stop solution (0.3 M sodium acetate). The reactions were purified by extraction with 200 μl phenol/chloroform and were precipitated in the presence of 1 μl of 20 mg/ml glycogen. The DNA was washed with 600 μl 70% ethanol and once the DNA pellets were dry they were resuspended in 8 μl of sequencing stop solution (0.3% each Bromophenol Blue and Xylene Cyanol FF, 10 mM EDTA pH 7.5 and 97.5% deionised formamide). After heating to 95°C for 2 minutes, 4 μl was run on a 6% denaturing gel at a constant temperature of 50°C (120W).

MODELLING OF THE HELICASE DOMAINS OF MFD

Amino acids 541-983 of *E. coli* Mfd were aligned using Clustal W with amino acids 309-755 of *Thermatoga maritima* RecG. A homology model was produced based on the coordinates of the *T. maritima* RecG-ADP-DNA complex [accession code 1GM5 (Singleton)] using InsightII 97 (Accelrys).

Images of protein structures were produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco.

CHAPTER 3

IN VIVO ROADBLOCK REPRESSION ASSAY

INTRODUCTION

The effect of DNA-bound proteins on transcription elongation

Lesions in the template strand of DNA are capable of stalling transcription complexes. Pyrimidine dimers (Selby *et al.*, 1997; Tornaletti *et al.*, 1999) and psoralen interstrand crosslinks (Shi *et al.*, 1988a; Shi *et al.*, 1988b) have been shown to arrest transcription elongation complexes. In addition malondialdehyde, 8-oxoG, abasic sites and gaps ((Kalogeraki *et al.*, 2003; Tornaletti *et al.*, 2004; Zhou and Doetsch, 1993), reviewed in (Tornaletti and Hanawalt, 1999) impede transcription to varying extents without necessarily completely halting the transcribing RNA polymerase.

Within a cell, the template DNA may also be bound by DNA-binding proteins, which may act as a physical "roadblock" to transcription and prevent progression of the transcribing RNAP. A mutant of the restriction enzyme *Eco*RI in which E111 has been substituted with either alanine or glutamine is defective in cleavage and therefore remains bound to its recognition site. This mutant blocked transcription by several different RNA polymerases *in vitro* on both circular and linear templates (Pavco and Steege, 1990). Transcription by *E. coli*, T7 and SP6 RNAPs was stalled by this *Eco*RI mutant, although greater read-through was observed for the viral polymerases (Pavco and Steege, 1991). The Lac repressor protein has also been shown to be capable of acting as a roadblock to transcription (Deuschle *et al.*, 1986). Placing a *lac* operator site between a promoter and a reporter gene prevented expression of the reporter in the presence of Lac repressor (Deuschle *et al.*, 1986). Truncated transcripts corresponding to the position of the Lac repressor were identified *in vitro* and *in vivo* and the blockage to transcription was removed by the

addition of IPTG (isopropyl-β-D-thiogalactoside). Lac repressor has also been shown to be capable of stalling eukaryotic RNAP II transcription (Reines and Mote, 1993). The addition of TFIIS caused transcript cleavage to permit chain re-elongation once the repressor dissociated. It should be noted that not all proteins bound to DNA act as roadblocks to transcription, for example TFIIIA bound to the internal control region of 5S rRNA is displaced by transcribing T7 RNA polymerase (Campbell and Setzer, 1991).

The majority of studies into the effect of DNA-bound proteins on transcription were performed *in vitro* and under single round transcription conditions. Multi-round transcription reactions were performed on templates blocked by a cleavage defective mutant of the *Eco*RI restriction enzyme and it was found that the presence of a second trailing RNAP caused read-through of the blockage *in vivo* and *in vitro* (Epshtein *et al.*, 2003). The second polymerase may cooperate in roadblock read-through by preventing the leading RNAP entering a backtracked conformation when it encounters the protein blockage (Epshtein and Nudler, 2003; Epshtein *et al.*, 2003). However, backtracking at an *Eco*RI EQ111 roadblock was not observed in single-round experiments where this mutant was used to probe the mechanism of transcription termination (Santangelo and Roberts, 2004).

Physiological role of roadblock repression

Regulated formation of protein roadblocks has been shown to be important in controlling the expression of some bacterial operons. The arabinose operon of *Bacillus subtilis* contains *cres* binding sites for the carbon catabolite repressor protein CcpA. The operon possesses two *cres* sites, one at +60 between the promoter and the *araA* gene and one two kilobases downstream within the *araB* gene. A construct containing *araA* fused to *lacZ* was repressed 5.5-fold in the

presence of glucose and repression was dependent on CcpA (Inacio *et al.*, 2003). The *B. subtilis dra-nupC-pdp* operon, which encodes proteins involved in the utilisation of deoxyribonucleotides, is also subject to catabolite repression by the CcpA protein. CcpA binds to a *cres* site 64 base pairs downstream of the transcription start point and blocks transcription of the downstream genes. In the presence of glucose, 5-fold repression was observed and repression occurred even at low frequencies of transcription initiation (Zeng *et al.*, 2000).

The *E. coli purB* operon has an operator site for the purine repressor, PurR, 242 bp downstream of the transcription start side overlapping codons 62 to 67 of the coding region. Binding of the repressor results in repression of the operon in a manner that is independent of promoter strength and coupled-translation and that is not affected by a 2.5-fold increase in operator-repressor affinity (He and Zalkin, 1992). Transcripts truncated at the operator site were observed *in vivo* in the presence of repressor although these truncated transcripts were not observed *in vitro*.

The mechanism of roadblock repression of the *E. coli bgl* operon is Rho-dependent (Dole *et al.*, 2004). H-NS (histone-like nucleoid structuring protein) bound 600 bp into the coding region of *bglG* causes transcription to stall. It is proposed that the stalling of transcription at the roadblock permits Rho to travel along the RNA and cause termination. This proposal is supported by the fact that repression is reduced from 7-fold to 3-fold during translation, presumably due to ribosomes preventing access of Rho to the RNA.

Binding of Lac repressor to an operator site placed downstream of a promoter led to the generation of truncated transcripts *in vivo* (Deuschle *et al.*, 1986). The Lac repressor has also been shown to lead to the formation of truncated transcripts *in vivo* at a Lac operator binding site located between the *lacI* and *ZYA* genes (Sellitti *et al.*,

1987). In addition, binding of Lac repressor to an operator within the 3′ region of the *lacI* gene led to the formation of truncated *lacI* transcripts that were subject to tagging by the SsrA system. The protein encoded by the tagged truncated transcripts was subsequently degraded by proteolysis (Abo *et al.*, 2000).

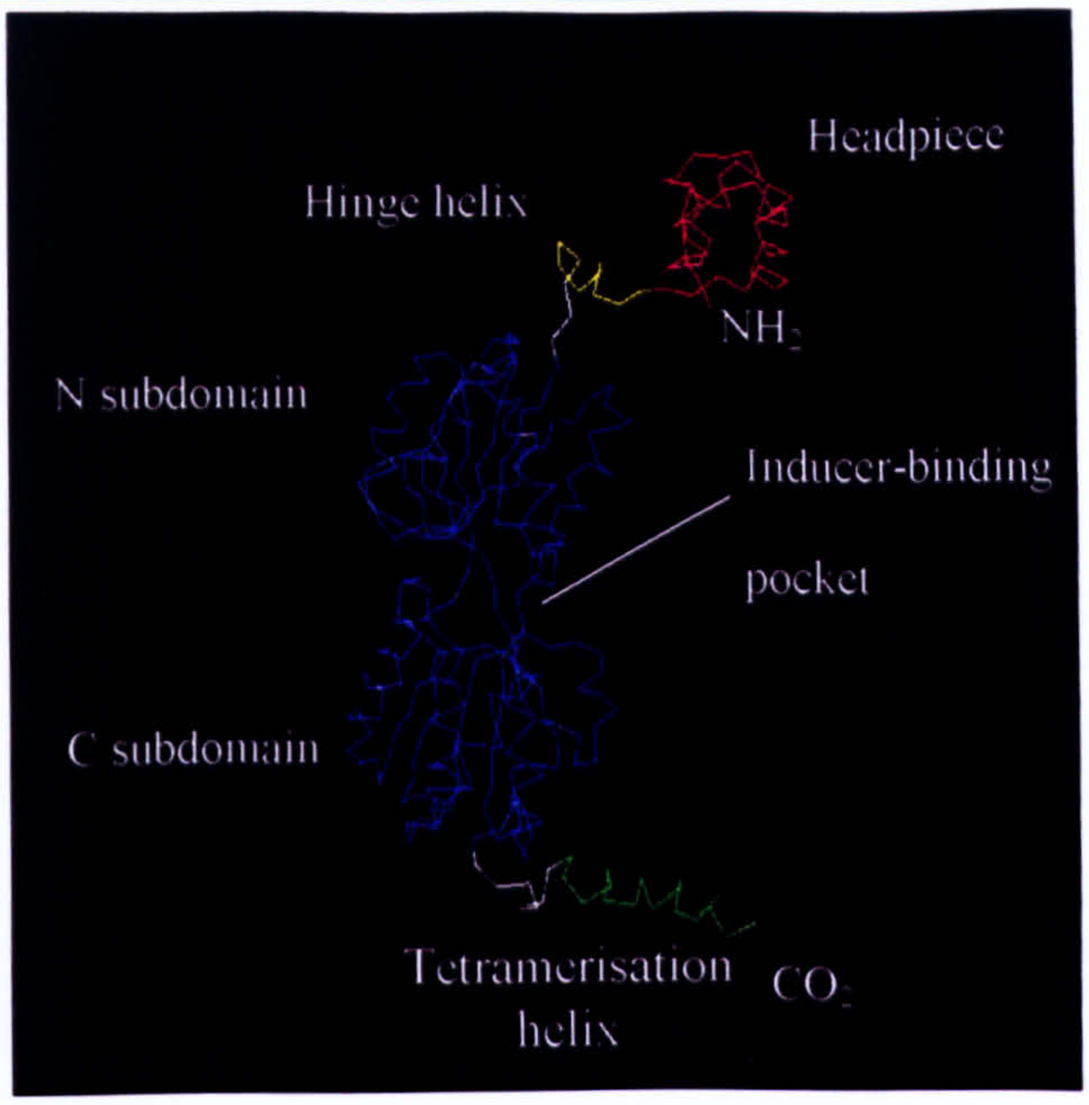
Lac repressor-operator interactions

The Lac repressor protein is the product of the *lacI* gene and forms a 154 kDa homotetramer. *In vivo* the repressor regulates expression of the *lac* operon i.e. *lacZ, lacY* and *lacA* genes, by binding to three *lac* operator sites. IPTG binds to the Lac repressor and reduces its affinity for the operator sites. Crystal structures (Friedman *et al.*, 1995; Lewis *et al.*, 1996; Slijper *et al.*, 1996) show that each Lac repressor monomer is composed of an N-terminal headpiece that is responsible for binding to DNA, a hinge region, a ligand-binding domain and a C-terminal helix that is responsible for oligomerisation via the formation of a four-helix bundle (figure 3.1 A).

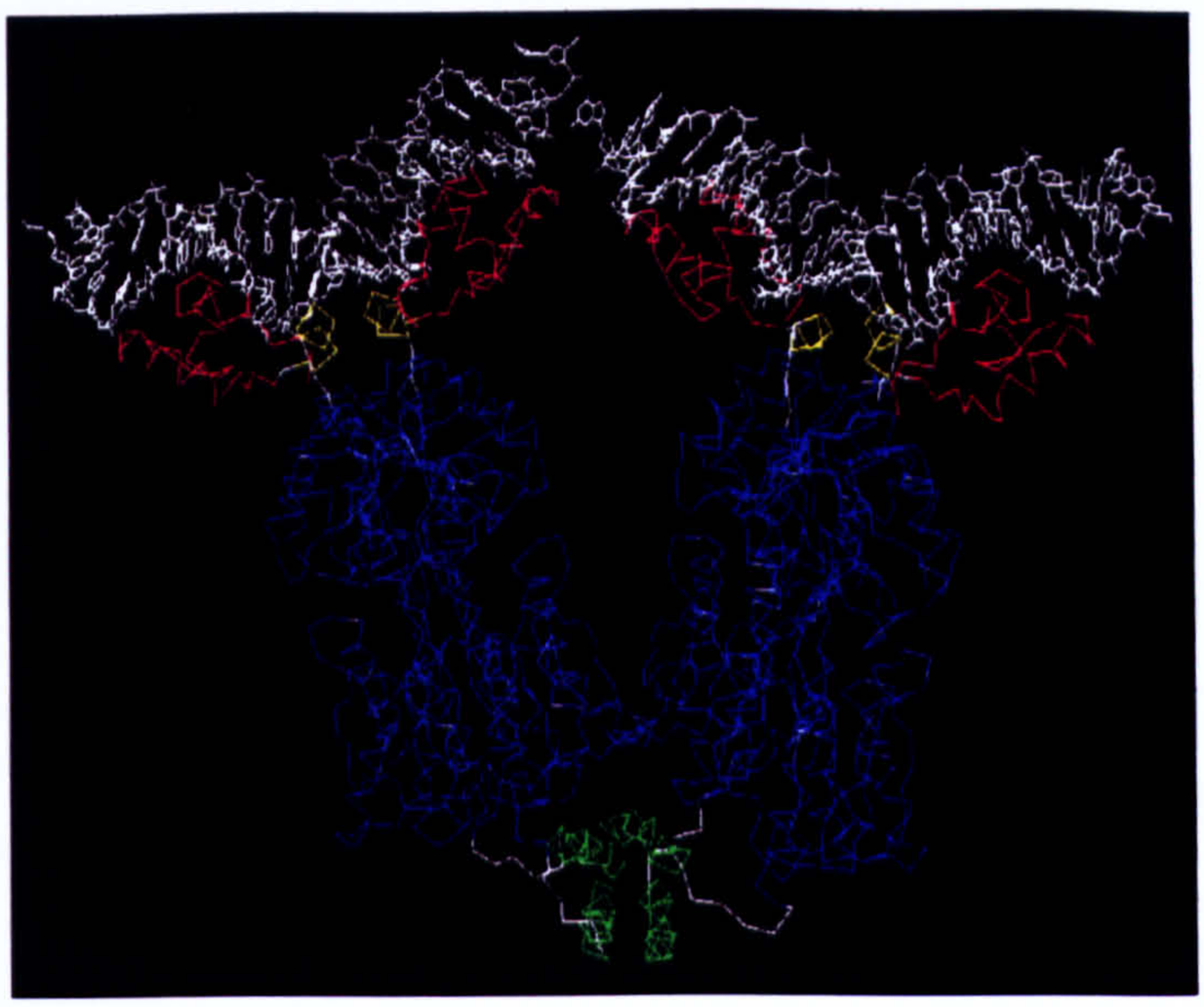
The DNA-binding domain contains a helix-turn-helix motif that recognises the operator DNA and is connected via the hinge to the core of the repressor. When bound to DNA the hinge region becomes ordered and forms an α-helix that inserts into the minor groove of the DNA. Binding of IPTG to the core domain disorders the hinge region, causing reorientation and increased flexibility of the N-terminal headpiece and therefore decreases DNA-binding affinity.

Both the fully symmetric 21 bp "ideal" operator DNA and a 22 bp natural operator *O1* sequence are bent 40° when bound to Lac repressor (Lewis 1996, Bell 2001). The "ideal" operator site is a perfectly symmetric sequence that is an inverted repeat of the left-half of *O1* with the central G.C pair deleted (figure 3.1). Lac repressor

A



B



C

Wild-type *O1* operator	5′ AATTGTGAGCGGATAACAATT 3′
O2 operator	5′ AAATGTGAGCGAGTAACAACC 3′
O3 operator	5′ GGCAGTGAGCGCAACGCAATT 3′
'Ideal' operator	5′ AATTGTGAGC-GCTCACAATT 3′

Figure 3.1. Lac repressor and operators

(A) Structure of Lac repressor monomer (Bell and Lewis, 2001). The DNA-binding headpiece is shown in red, the hinge helix in yellow, the core in blue and the tetramerisation helix in green. (B) Structure of Lac repressor tetramer bound to operator DNA (Bell and Lewis, 2001). (C) Sequence of natural *lac* operators. The central G:C base pair is shown in blue and bases in *O2* and *O3* that do not match *O1* are shown in red. The 'ideal' operator is a perfect inverted repeat of the left half of the *O1* operator lacking the central G:C pair.

binds the synthetic operator 8-fold tighter than wild-type operator sequence (Simons *et al.*, 1984) . In addition to the *lac O1* operator the *lac* promoter also contains two pseudo-operators, *O2* and *O3*. The destruction of just one of these pseudo-operator sites was shown to decrease repression by Lac repressor only 2-fold but when both pseudo-operator sites were absent, repression decreased 70-fold (Oehler *et al.*, 1990). This indicated that cooperative interaction was possible between operators through DNA loop formation mediated by the tetrameric Lac repressor.

The phase dependence of DNA loop formation between *lac O1* and an auxiliary operator was investigated at spacings between 57.5 and 1493.5 bp (Muller *et al.*, 1996). Maximal repression occurred with periodicity and repression was increased at short inter-operator distances. The presence of an auxiliary operator served to increase the occupancy of *lac O1*. With Lac repressor present at around 10-15 copies per cell, the binding of one dimer of a tetramer to an auxiliary operator allows the other dimer to search for a nearby operator thus increasing the local Lac repressor concentration for *O1* and as a consequence increasing *lac O1* occupancy. Stable loop formation caused maximal repression at a spacing of 70.5 bp with further maxima at 92.5 and 115.5 bp; these match the spacings that occur naturally within the *lac* operon.

A mutant of Lac repressor, which is capable of dimerisation but not tetramerisation, repressed the wild-type *lac* promoter to an extent that was comparable to the level of repression observed when both of the auxiliary operators were destroyed (Muller *et al.*, 1996; Muller-Hill, 1998). Some residual periodicity of maximal repression remained with this mutant, presumably due to non-specific interaction of the free C-terminal oligomerisation domain with the DNA. Loops formed by the Lac repressor have been visualised in minicircles by atomic force microscopy (Virnik *et*

al., 2003) and electron microscopy (Kramer *et al.*, 1988), and loops formed using Lac repressor protein have been used to create insulator-like elements that disconnect an enhancer from its promoter by placing them on separate DNA loops (Bondarenko *et al.*, 2003).

The role of Mfd in roadblock repression

Although repression of the *bgl* operon of *E. coli* was dependent on Rho, this is not true for all roadblock repression. The Mfd protein has been implicated in roadblock repression as its absence in *mfd*⁻ cells was shown to cause the partial relief of repression of two operons. Firstly, the *dra-nupC-pdp* operon, mentioned above, displays 5-fold catabolite repression in *mfd*⁺ cells but only 2.5-fold catabolite repression in *mfd*⁻ cells (Zeng *et al.*, 2000). This is equivalent to a 2-fold contribution of Mfd to roadblock repression. Secondly, a null mutation was discovered within the *mfd* gene that relieved catabolite repression of the *hut* and *gnt* operons in *Bacillus subtilis* (Zalieckas *et al.*, 1998). Catabolite repression in these operons is dependent on the binding of the CcpA repressor to *cre* sites. A 13-fold catabolite repression effect was observed in wild-type cells but this was reduced to 4.5-fold in cells containing the mutation in the *mfd* gene. This is equivalent to a 2.9-fold Mfd effect. The effect of Mfd was only observed at roadblocks downstream of the promoter; repression of the *amyE* and *bglPH* genes that contain *cre* sites within their promoters was not affected by the *mfd* mutation.

The Mfd protein has been shown to displace elongation complexes stalled at roadblocks *in vitro* (Selby and Sancar, 1995a). Based on the fact that catabolite repression was partially relieved in *mfd*⁻ cells, it was proposed that Mfd is able to displace transcription elongation complexes stalled at protein roadblocks *in vivo*. Creation of an *in vivo* roadblock repression assay would allow examination of

whether the effects of Mfd on roadblock repression *in vivo* are dependent on its ability to displace RNAP and therefore whether the effect of Mfd on roadblock repression is direct or indirect.

Null-mutations in the *mfd* gene caused only a slight increase (3-fold) in UV sensitivity (Selby and Sancar, 1993a). Therefore the natural phenotype of *mfd*⁻ cells is not readily assayed and so is not suitable for screening for mutants with disrupted *mfd* function. Therefore, development of an *in vivo* assay of Mfd function with a readily assayable phenotype would provide a means to isolate mutants that allow the function of Mfd to be dissected and, if the effect of Mfd on roadblock repression is dependent on its ability to displace ECs, would provide an *in vivo* assay for RNAP displacement.

RESULTS

Rationale behind roadblock reporter construction

It has been proposed that the effect of Mfd on roadblock repression is due to its ability to displace elongation complexes stalled at protein roadblocks. An *in vivo* roadblock repression assay was developed in order to examine this hypothesis in *E. coli* and to provide a phenotype for *mfd* cells that is readily assayed and therefore can be used in screening to isolate mutants of *mfd* that are defective in Mfd function. A roadblock repression reporter plasmid (pRCB-CAT) was constructed. The *cat* reporter gene was transcribed from the constitutive *galPcon6* promoter (Burns and Minchin, 1994). CAT expression was regulated by Lac repressor bound to an operator located between the promoter and the *cat* gene. To increase the occupancy of this operator, a second operator was located downstream of the *cat* gene and in order to obtain a high level of repression, two ideal *lac* operator sites (Simons *et al.*, 1984) were used. The centre-to-centre distance between the two operator sites was 815 bp. At this spacing, a Lac repressor tetramer bound to the downstream operator would be expected to increase the occupancy of the upstream operator 2-fold by looping of the intervening DNA (Muller *et al.*, 1996). The operators were placed either side of the *cat* gene so that only one roadblock was formed between the promoter and the reporter gene. Placing the second operator upstream of the promoter could potentially have interfered with transcription initiation by placing the promoter in a loop formed between the two operator sites.

If Mfd increases the efficiency of roadblock repression by removing RNAP complexes that have stalled at the roadblock site, a difference in the level of transcription of the *cat* reporter, in the presence or absence of Mfd, should be

observed. In the wild-type situation, where Mfd is present, the reporter activity should be low since stalled complexes will be displaced. However, in *mfd*$^-$ cells, or in cells containing an Mfd protein that is defective in displacement of RNAP, the stalled elongation complex will remain bound at the roadblock until the Lac repressor (in a binding equilibrium with its operator) dissociates from the DNA. Since the blockage to transcription is no longer present the RNAP can resume elongation and transcribe through the *cat* gene, resulting in increased reporter activity (figure 3.2).

The effect of Mfd on the efficiency of roadblock repression

Before the effect of *mfd* on roadblock repression could be considered, it needed to be determined that expression of the *cat* reporter gene in the roadblock repression construct pRCB-CAT was decreased by binding of Lac repressor to the operator. UNCNOMFD (*mfd*$^-$) cells were transformed with the roadblock reporter construct pRCB-CAT and the empty vector pET21a, which carries the *lacI*^{q} gene, encoding the Lac repressor. Cultures were grown to mid-exponential phase in the absence (repressed state) or presence (induced state) of IPTG and the CAT activities were measured (figure 3.3). The CAT activity in the repressed state (conditions in which the Lac repressor should be bound to operator DNA) was 9.5-fold lower than in the induced state (when the Lac repressor should not be DNA-bound). Therefore in this system Lac repressor bound to operator DNA repressed transcription of the *cat* gene 9.5-fold, presumably by acting as a block to transcription elongation.

Previous reports have shown that Mfd is able to increase the efficiency of roadblock repression of some bacterial operons *in vivo*. In order to be used as an *in vivo* assay

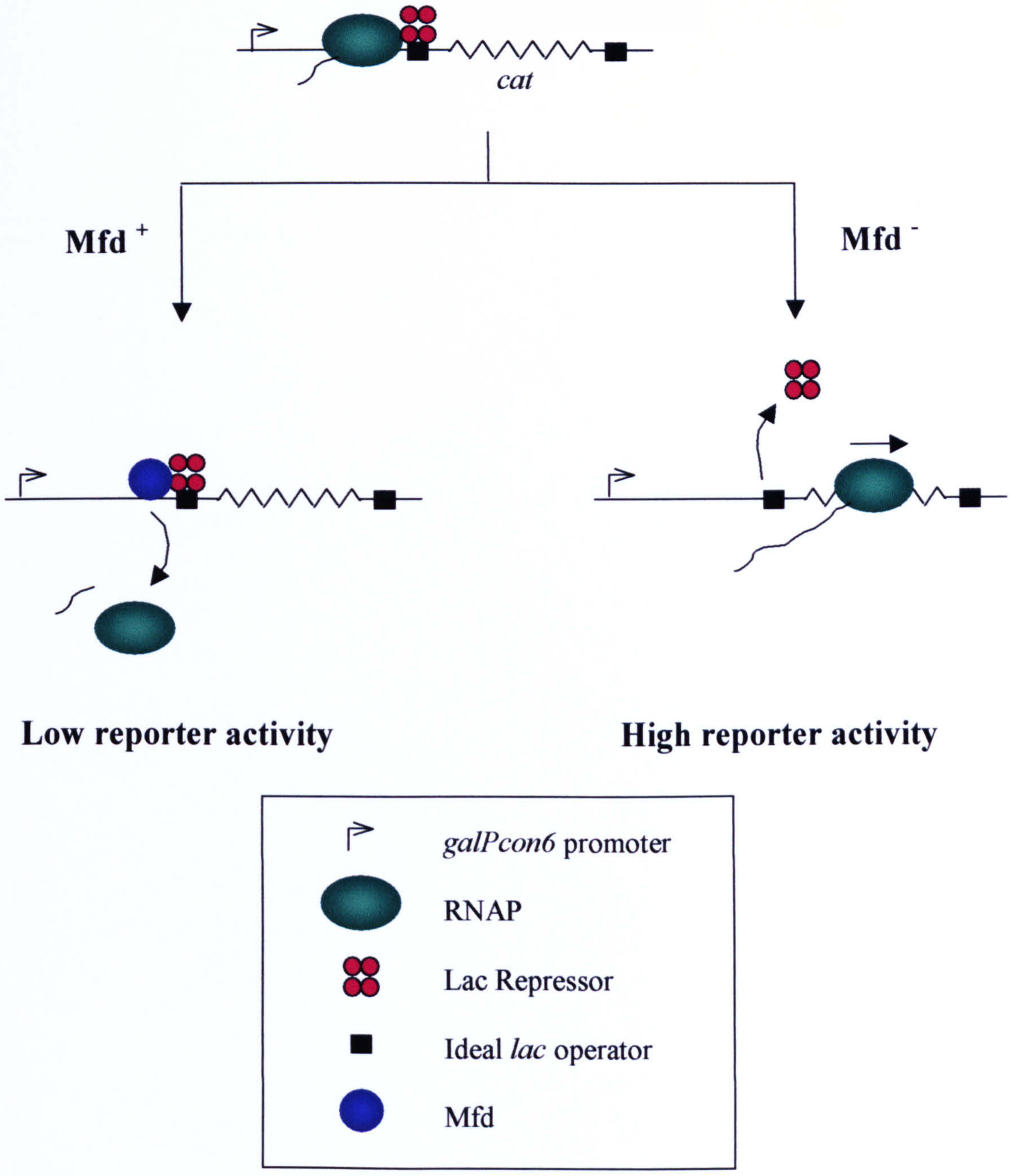
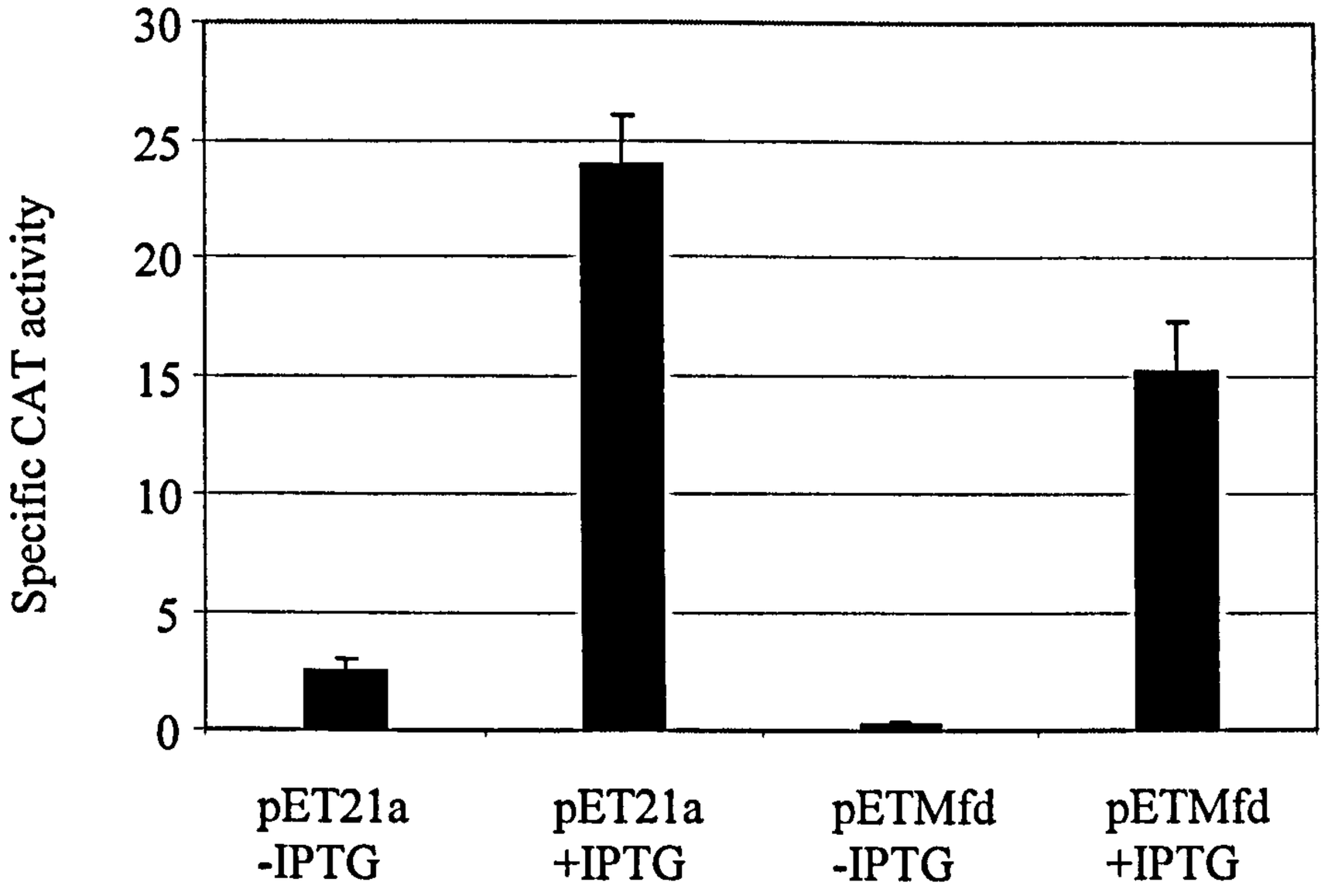


Figure 3.2. Roadblock reporter assay.

Transcription of a *cat* reporter gene from a *galPcon6* promoter is repressed by the binding of Lac repressor to an operator site upstream of the reporter gene. If Mfd contributes to roadblock repression, elongation complexes stalled by the protein roadblock will be displaced in the presence of Mfd, resulting in low levels of transcription of *cat*. In the absence of Mfd, the elongation complex will remain stalled at the roadblock until the repressor dissociates, at which point transcription can resume, resulting in higher levels of *cat* transcription.



Mfd	+IPTG (induced)	-IPTG (repressed)	Repression (+IPTG/-IPTG)
None	24.0 ± 2.1	2.5 ± 0.5	9.5-fold
Wild-type	15.2 ± 2.0	0.3 ± 0.1	61-fold

Figure 3.3. Roadblock repression of pRCB-CAT by Lac repressor *in vivo*.

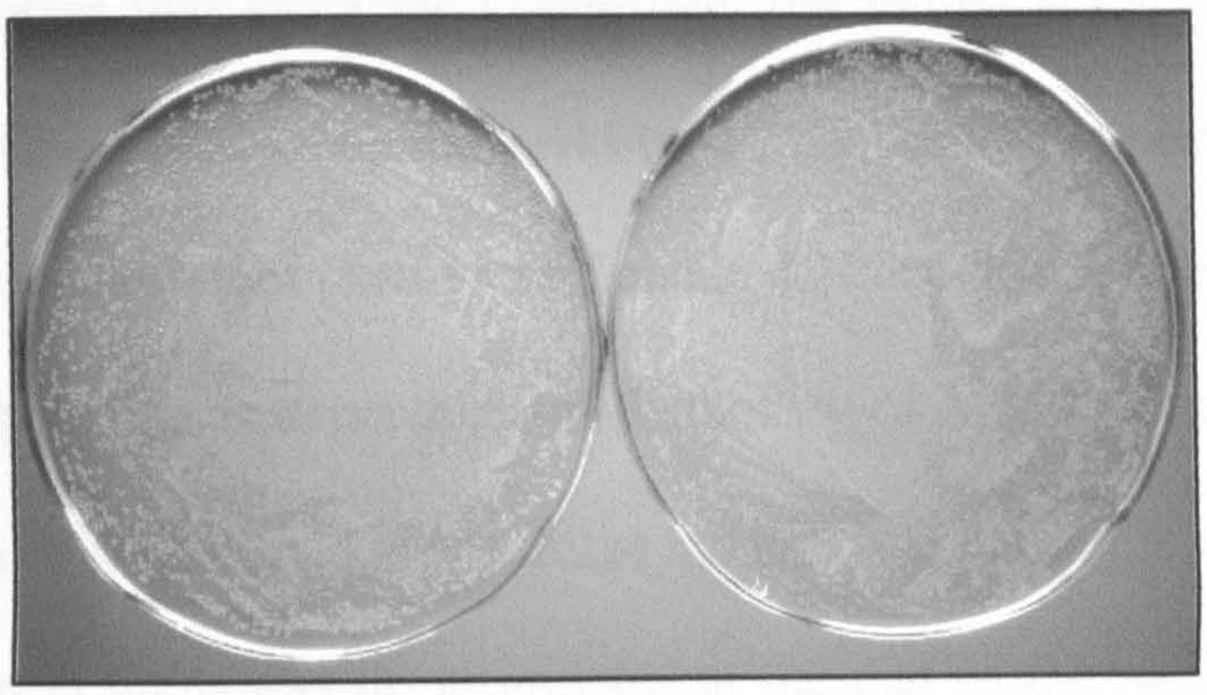
UNCNOMFD (*mfd*) cells containing the roadblock repression reporter pRCB-CAT were transformed with the plasmid indicated. pET21a was used as a control plasmid since it does not express Mfd. Cultures were grown in M9 media containing the appropriate antibiotics and 0.5 mM IPTG where indicated. Cells were harvested at OD_{600} ~0.5 and CAT assays were performed. Values are the average of three independent experiments and are shown with s.d. Specific CAT activity units are nmol chloramphenicol acetylated/min/mg protein.

for Mfd function in *E. coli* Mfd should enhance roadblock repression of the *cat* gene in the reporter construct pRCB-CAT.

UNCNOMFD (*mfd*) cells were transformed with the roadblock reporter construct pRCB-CAT and either pETMfd (expressing Mfd from its own promoter) or the empty vector pET21a as described previously. Like pET21a, the pETMfd plasmid carries the *lacI*^{q} gene. Cultures were grown to mid-exponential phase in the absence (repressed state) or presence of IPTG (induced state) and the CAT activity was determined (figure 3.3). In the presence of Mfd the CAT activity in the repressed state was 61-fold lower than in the induced state. Although the presence of Mfd resulted in a decrease in CAT activity in both the repressed and induced states the effect was much more pronounced in the repressed state. In these experiments Mfd expression had an 8-fold effect on CAT activity in the repressed state. These data show that Mfd increases the efficiency of repression of a Lac repressor roadblock on the reporter construct pRCB-CAT in *E. coli*.

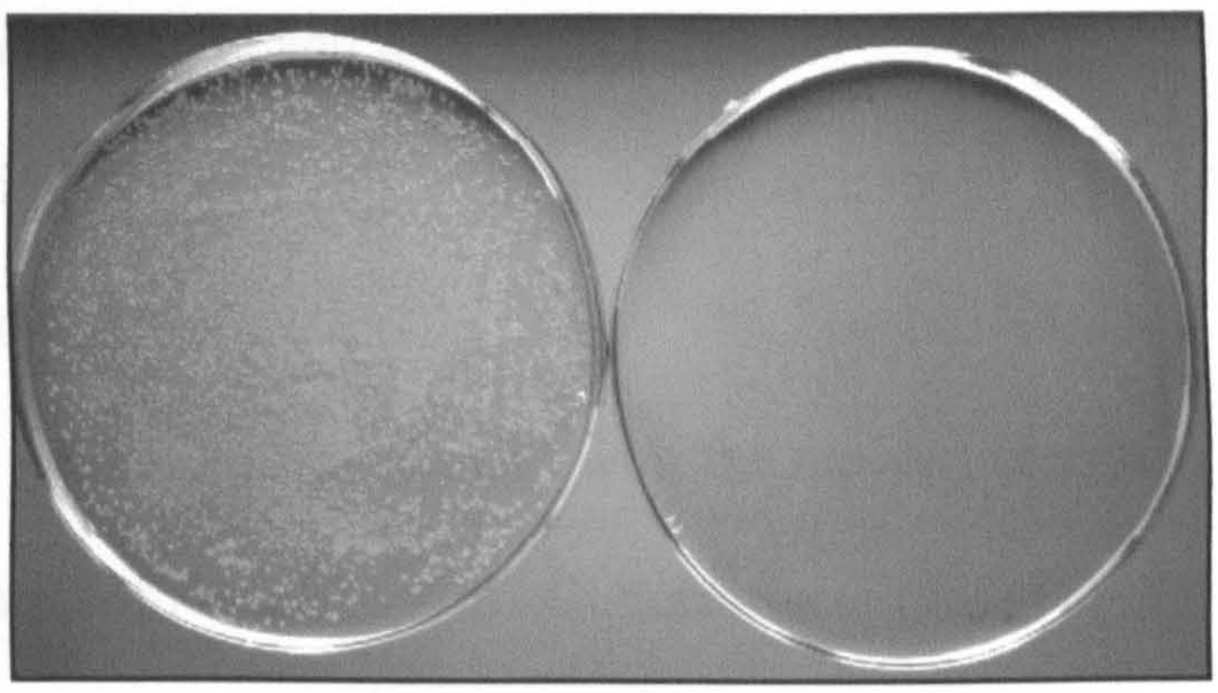
To determine whether the difference in CAT activity observed between cells lacking Mfd and containing functional Mfd was sufficient for differential growth on agar plates containing chloramphenicol, UNCNOMFD (*mfd*) cells containing the reporter pRCB-CAT were transformed with pET21a or pETMfd and were spread onto plates containing 5 μg/ml chloramphenicol. The cells containing pET21a grew readily but when the same cells were transformed with pETMfd, expressing wild-type Mfd, the bacteria were chloramphenicol-sensitive and no growth was observed on plates containing 5 μg/ml chloramphenicol (figure 3.4). The 8-fold difference in CAT activity between cells expressing functional Mfd and those lacking Mfd was sufficient to result in a chloramphenicol-resistant phenotype for cells lacking Mfd.

pET21a



Chloramphenicol 0 μg/ml 5 μg/ml

pETMfd



Chloramphenicol 0 μg/ml 5 μg/ml

Figure 3.4. The chloramphenicol resistance phenotype of *mfd*⁻ cells containing the reporter construct pRCB-CAT.

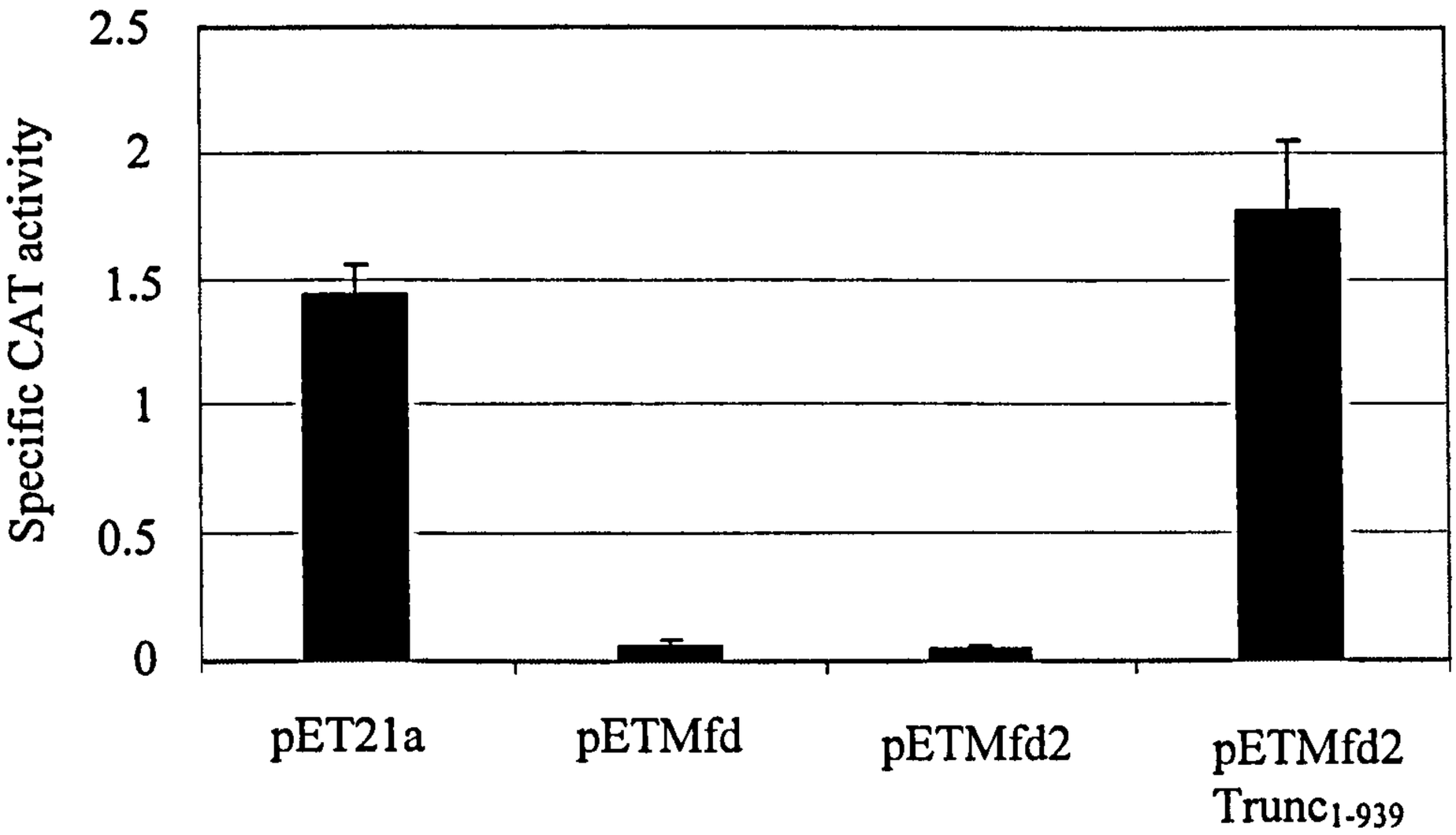
UNCNOMFD cells were transformed with 20 ng pET21a plasmid or 20 ng pETMfd plasmid. The transformations were spread onto M9 plates containing kanamycin, tetracycline and ampicillin and where indicated, 5 μg/ml chloramphenicol.

The effect of Mfd_{trunc} on roadblock repression

A truncated Mfd protein that lacks the C-terminal 209 amino acids (Mfd_{trunc}) has been shown to bind RNAP and DNA like wild-type Mfd but has a specific defect in elongation complex displacement (Selby and Sancar, 1995a). If the effect of Mfd on *in vivo* roadblock repression is dependent on the ability of Mfd to displace RNAP, the truncated version of Mfd should not increase the efficiency of repression.

In the pETMfd plasmid there was no terminator at the end of the *mfd* gene and as such there was a possibility that transcription could continue into the adjacent $lacI^q$ gene and affect roadblock repression. Therefore before truncation of Mfd was tested, the plasmid pETMfd2 was constructed, which contains the *mfd* gene under the control of its own promoter and also carries the $lacI^q$ gene. The *mfd* gene is in the reverse orientation to pETMfd and a λoop terminator has been inserted downstream of the stop codon.

UNCNOMFD (*mfd*$^-$) cells were transformed with the reporter construct pRCB-CAT and either pET21a, pETMfd or pETMfd2. Cultures were grown to mid-exponential phase and their CAT activity determined (figure 3.5). The introduction of a terminator and the change in orientation of the *mfd* gene in pETMfd2 compared to pETMfd did not alter the efficiency of roadblock repression. In comparison to cells containing pET21a repressed CAT activity was reduced 28-fold in cells expressing Mfd from pETMfd2 and 22-fold in cells containing pETMfd. This indicates that the lack of a terminator at the end of the *mfd* gene in pETMfd and therefore the possibility of continued transcription into the $lacI^q$ gene did not affect the ability of Mfd to increase roadblock repression efficiency.



Mfd Expression Plasmid	Repressed CAT activity	Mfd effect (-Mfd/+Mfd) in repressed state
pETMfd	0.06±0.01	23-fold
pETMfd2	0.05±0.01	28-fold
pETMfd2Trunc$_{1\text{-}939}$	1.8±0.3	0.8-fold

Figure 3.5. The effect of Mfd and a truncation mutant defective in RNAP displacement on roadblock repression in *mfd*$^-$ cells.

UNCNOMFD (*mfd*$^-$) cells containing the roadblock repression reporter pRCB-CAT were transformed with the plasmid indicated. pET21a was used as a control plasmid not expressing Mfd. Cultures were grown in M9 media containing the appropriate antibiotics. Cells were harvested at OD_{600} ~0.5 and CAT activity determined. Values are the average of three independent experiments and are shown with s.d. Specific CAT activity units are nmol chloramphenicol acetylated/min/mg protein.

Mfd protein truncated following amino acid 939 has been shown to be defective in displacement of stalled elongation complexes *in vitro* (Selby and Sancar, 1995a). The dependence on RNAP displacement activity of Mfd enhancement of roadblock repression can be tested using this truncation.

A derivative of pETMfd2, pETMfd2Trunc$_{1\text{-}939}$, was constructed that expresses Mfd truncated by a stop codon following amino acid 939. UNCNOMFD (*mfd*) cells containing the CAT reporter plasmid pRCB-CAT were transformed with pETMfd2Trunc$_{1\text{-}939}$. Cultures were grown to mid-exponential phase and the CAT activity measured (figure 3.5). Cells expressing Mfd truncated after amino acid 939 possessed the same level of CAT activity as cells that lacked Mfd.

As truncation of Mfd after amino acid 939 abolishes its ability to displace elongation complexes, these data indicate that the RNAP displacement activity of Mfd is required to increase roadblock repression efficiency *in vivo*. Therefore the *in vivo* roadblock repression system that has been developed provides an assay for Mfd mutants that are defective in elongation complex displacement. Mfd mutants that are defective in increasing roadblock repression are chloramphenicol-resistant and therefore growth on chloramphenicol plates provides an easy means of selecting for Mfd mutants that are defective in RNAP dissociation.

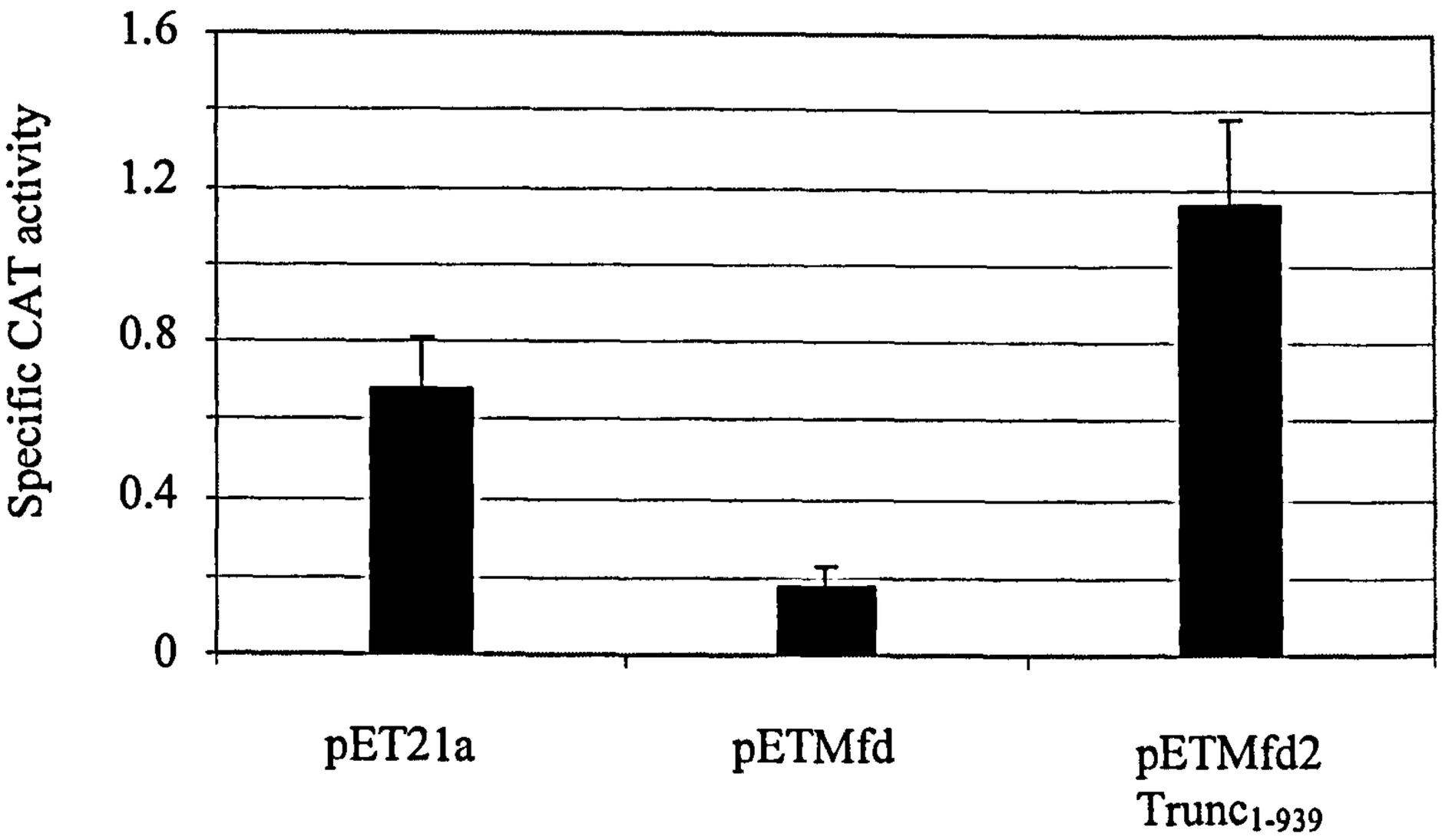
MfdTrunc$_{1\text{-}939}$ is still able to bind to RNAP and DNA but is unable to displace transcription elongation complexes (Selby and Sancar, 1995a). Therefore one might predict that its over-expression in cells expressing wild-type Mfd from the chromosome would have a trans-dominant effect i.e. its expression would inhibit the ability of the endogenous wild-type Mfd to dissociate elongation complexes. Binding of the truncated Mfd to stalled RNAP complexes would prevent binding of

wild-type Mfd and would therefore impair dissociation of stalled transcription complexes.

To determine whether MfdTrunc$_{1\text{-}939}$ has a trans-dominant effect on roadblock repression *in vivo*, AB1157 cells (with an intact chromosomal *mfd* gene) were transformed with the roadblock reporter plasmid pRCB-CAT and either pET21a, pETMfd or pETMfd2Trunc$_{1\text{-}939}$. Cultures were grown to mid-exponential phase and the CAT activity determined (figure 3.6).

Repressed CAT activity was 1.7-fold higher in *mfd*$^+$ cells over-expressing MfdTrunc$_{1\text{-}939}$ than in cells expressing only wild-type Mfd from the chromosome. This indicates that over-expression of mutant Mfd that cannot displace transcription elongation complexes interferes with the activity of wild-type Mfd in roadblock repression, presumably by competing with the wild-type protein for binding to stalled RNAP, but failing to remove the blocked complex.

The over-expression of wild-type Mfd from pETMfd in *mfd*$^+$ cells resulted in ~3-fold lower CAT activity than *mfd*$^+$ cells containing a control plasmid that did not express Mfd (figure 3.6), suggesting that wild-type levels of Mfd expression are sub-saturating.



Mfd Expression Plasmid	Repressed CAT activity	Mfd effect (-Mfd plasmid/+Mfd plasmid) in repressed state
pETMfd	0.2±0.05	3.5-fold
pETMfd2Trunc$_{1\text{-}939}$	1.2±0.2	0.58-fold

Figure 3.6. The effect of truncation of Mfd after amino acid 939 on roadblock repression in *mfd*$^+$ cells.

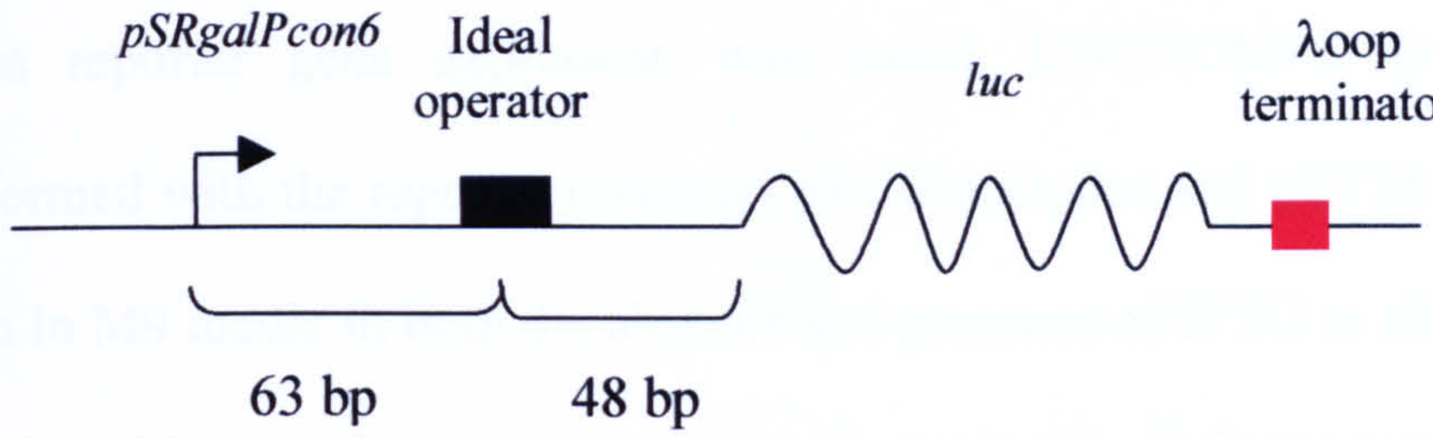
AB1157 (*mfd*$^+$) cells containing the roadblock repression reporter pRCB-CAT were transformed with the plasmid indicated. pET21a was used as a control plasmid to examine endogenous Mfd and pETMfd or pETMfd2Trunc$_{1\text{-}939}$ was used to over-express Mfd protein. Cultures were grown in M9 media containing the appropriate antibiotics. Cells were harvested at OD_{600} ~0.5 and CAT assays were performed. Values are the average of three independent experiments and are shown with s.d. Specific CAT activity units are nmol chloramphenicol acetylated/min/mg protein.

Luciferase roadblock repression system

Unfortunately after using using the *cat* roadblock assay system for 18 months, the assay kit used to quantify *cat* reporter activity was discontinued by the manufacturer. The use of an alternative assay for CAT activity such as thin layer chromatography and RT-PCR was considered but it was decided that these were not sufficiently sensitive and were not suited to the analysis of a large number of samples. Although the *cat* system was still able to be used for isolation of mutants of Mfd defective in RNAP displacement on plates containing 5 μg/ml chloramphenicol, a new roadblock reporter construct was required containing a different reporter gene in order to quantitatively assay effects on roadblock repression in liquid culture.

A new roadblock reporter plasmid was constructed using firefly luciferase as a reporter gene (figure 3.7). The rationale behind the design of the reporter construct was the same as for the *cat* reporter. The *luc* gene from pGL2, with an identical Shine-Dalgarno sequence to the *cat* roadblock reporter, was placed under the control of the *galPcon6* promoter. The luciferase reporter plasmid, pRCBlacO$_{ID}$luc, contained an 'ideal' *lac* operator 48 bp upstream of the start of the *luc* gene. Initially a second operator was not introduced since an auxiliary operator placed downstream of the 1.65 kb long luciferase gene would be too far away to be expected to increase occupancy of the upstream operator located between the promoter and the reporter gene.

pRCBlacO$_{ID}$luc



pRCBKA4

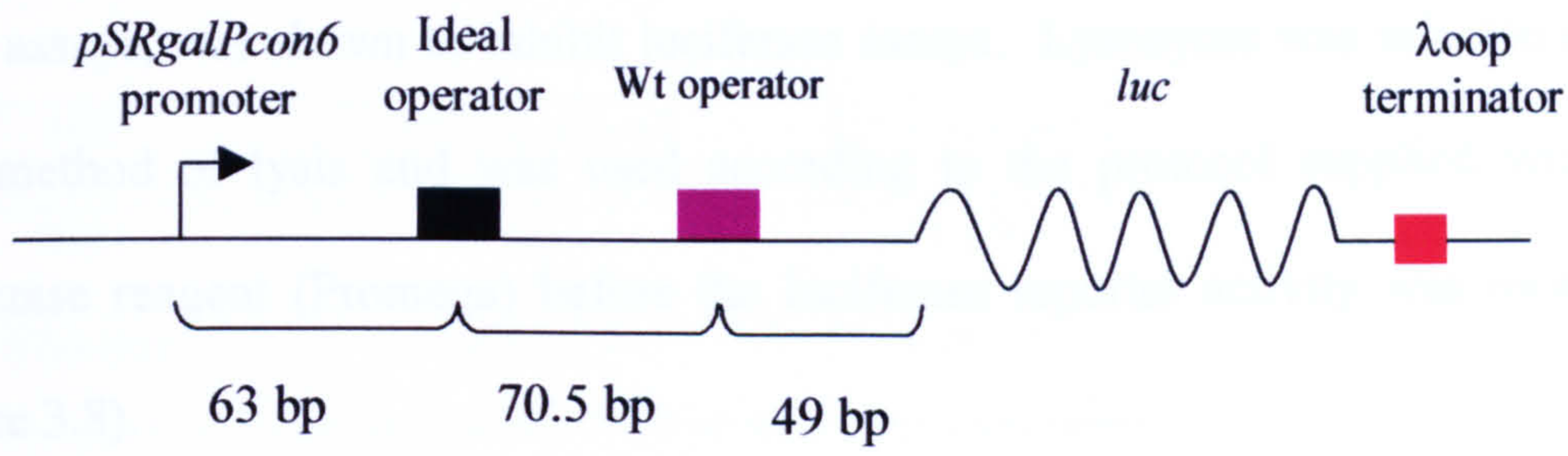


Figure 3.7. Luciferase roadblock reporter constructs.

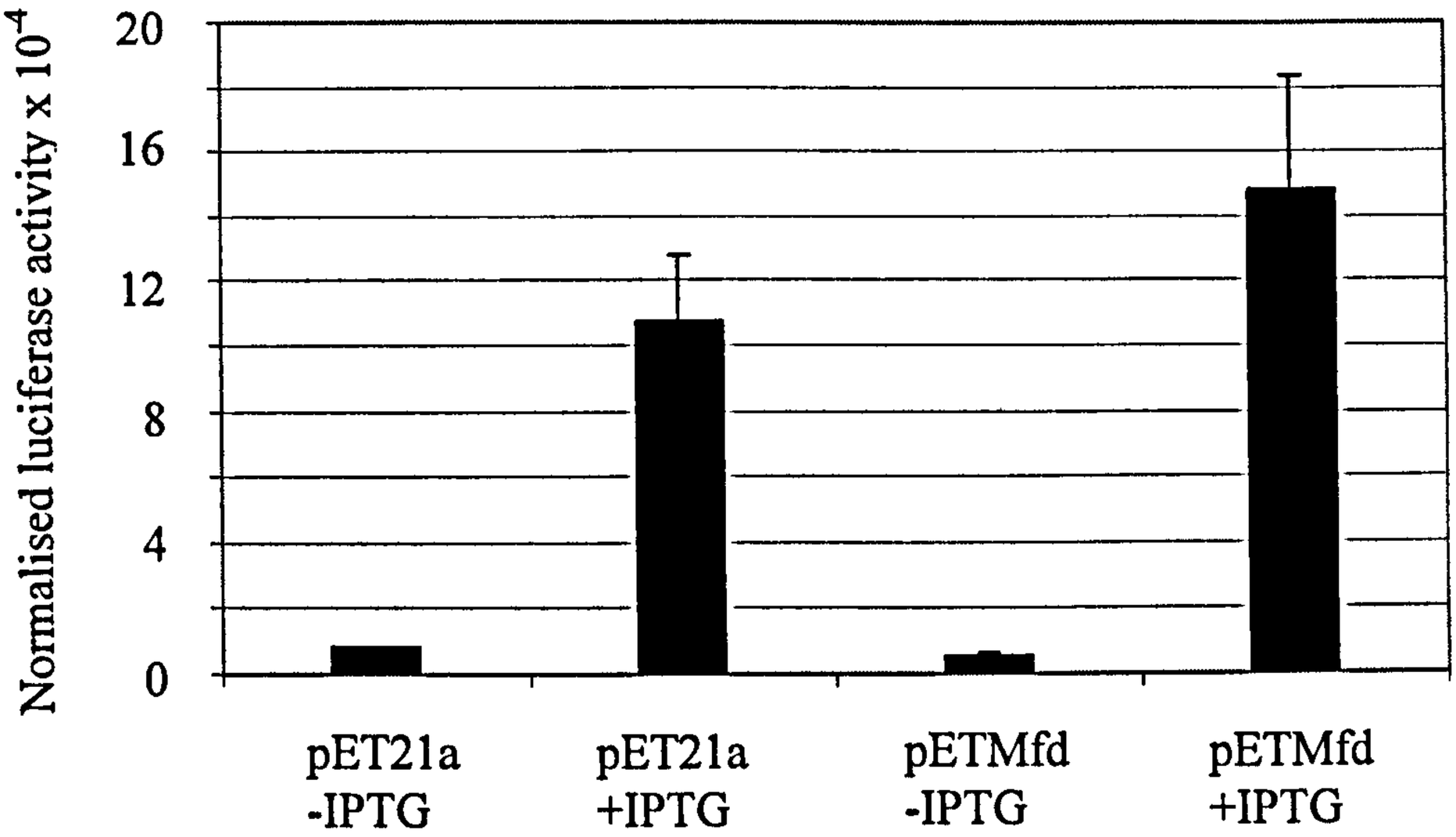
pRCBlacO$_{ID}$luc contains an ideal *lac* operator site 48 bp upstream of the firefly *luc* gene. Transcription of the reporter gene initiates at a constitutive *pSRgalPcon6* promoter and terminates at a λoop terminator. pRCBKA4 contains a wild-type *lac* operator site 49 bp upstream of the *luc* gene and an 'ideal' *lac* operator a further 70.5 bp upstream.

Roadblock repression by a single operator in luciferase system

To ensure comparability of the luciferase reporter system with the previous *cat* system, key experiments were repeated using the new reporter. Firstly the ability of a single ideal *lac* operator placed between the promoter and luciferase gene to repress reporter gene expression was tested. UNCNOMFD (*mfd*) cells were transformed with the reporter construct pRCBlacO_{ID}luc and pET21a. Cultures were grown in M9 media in both the absence and presence of IPTG to allow the repressed and induced levels of gene expression to be measured. Cultures were grown to mid-exponential phase and the cells were harvested and lysed. A variety of lysis methods were tested and notably Bugbuster (Novagen), which had been used to lyse cells for CAT assays, was shown to inhibit luciferase assays. Lysozyme was selected as the best method of lysis and was used according to the protocol supplied with the luciferase reagent (Promega) before the luciferase reporter activity was measured (figure 3.8).

Luciferase activity was 13-fold lower under repressed conditions than under induced conditions, indicating that the binding of Lac repressor to the ideal *lac* operator site between the promoter and the luciferase gene was sufficient to reduce transcription of the downstream reporter gene.

The displacement activity of the Mfd protein was required for increased roadblock repression of the previous *cat* reporter system. To determine whether the new luciferase reporter system was also dependent on the ability of Mfd to displace stalled elongation complexes, UNCNOMFD cells containing pRCBlacO_{ID}luc were transformed with the control plasmid pET21a, pETMfd or pETMfd2Trunc$_{1\text{-}939}$. Cultures were grown to mid-exponential phase and the reporter activity in the



Mfd	+IPTG (induced)	-IPTG (repressed)	Repression (+IPTG/-IPTG)
None	10.7 ± 2.1	0.8 ± 0.1	13.4-fold
Wild-type	14.8 ± 3.5	0.5± 0.1	29.6-fold

Figure 3.8. Repression of pRCBlacO$_{ID}$luc by Lac repressor.

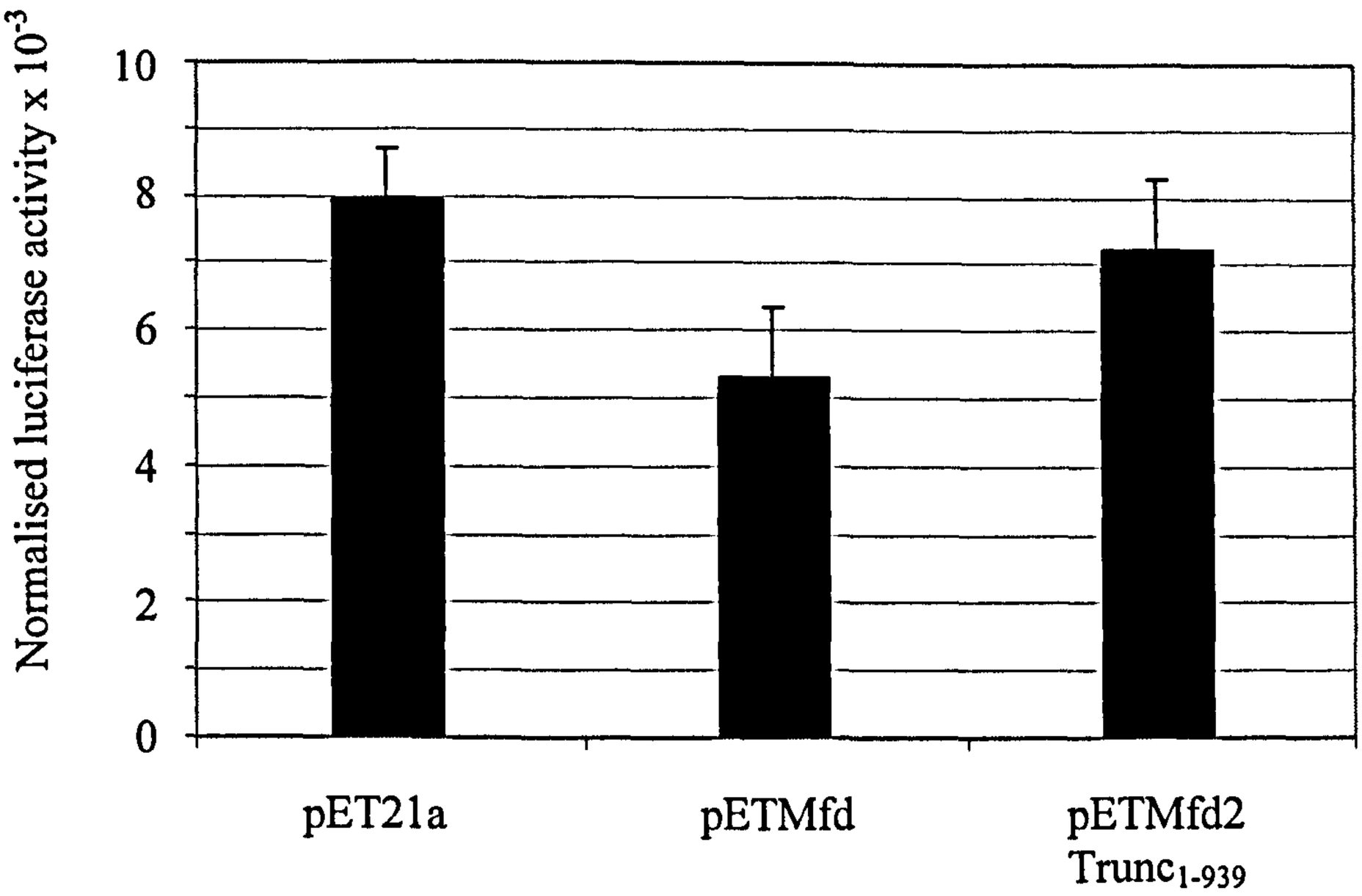
UNCNOMFD (*mfd*) cells containing the reporter construct pRCBlacO$_{ID}$luc were transformed with either pETMfd or pET21a. Cultures were grown to mid-exponential phase (OD_{600} ~0.5) in M9 media containing the appropriate antibiotics and where indicated 0.5 mM IPTG. Luciferase activity was measured and the normalised luciferase activity calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested. The values shown are an average of three independent experiments and are shown with s.d.

repressed state was assayed (figures 3.8 and 3.9). The expression of Mfd resulted in only a 1.6-fold reduction of luciferase activity and therefore did not increase roadblock repression at the single operator of the luciferase reporter to the same extent (8-fold) as had been observed for the pRCB-CAT reporter. The C-terminal truncation of Mfd, which lacks RNAP displacement activity, did not reduce luciferase activity significantly in comparison to cells lacking the Mfd protein.

Whilst the efficiency of roadblock repression of the *cat* gene in the pRCB-CAT reporter construct was increased by Mfd in a manner dependent on its ability to displace stalled elongation complexes, the effect of Mfd on expression of the luciferase gene was only modest. Therefore pRCBlacO$_{ID}$luc is not a suitable roadblock reporter construct to assay Mfd function *in vivo*.

Roadblock repression by tandem operators in luciferase system

The luciferase reporter construct pRCBlacO$_{ID}$luc, with a single *lac* operator site, did not provide a suitable assay for *in vivo* Mfd function. Therefore a second luciferase reporter construct was designed containing two *lac* operators. As mentioned previously, the luciferase gene is too long for an auxiliary operator placed downstream of the reporter gene to increase the occupancy of the operator located between the promoter and the reporter gene. Therefore the second luciferase roadblock reporter, pRCBKA4, contained one wild-type *lac* operator 49 bp upstream of the *luc* gene and a second 'ideal' *lac* operator a further 70.5 bp upstream (centre-to-centre distance) (figure 3.7). This operator spacing was selected since it had been shown to result in the greatest repression, 50-fold greater than a single operator, of a series of operator spacings tested by Müller (Muller *et al.*, 1996). Once again, in order to determine whether the system was comparable to the *cat* reporter system,



Mfd Expression Plasmid	Repressed normalised luc activity x10^{-3}	Mfd effect (-Mfd/+Mfd) in repressed state
pETMfd	0.5±0.1	1.6-fold
pETMfd2Trunc$_{1-939}$	0.7±0.1	1.1-fold

Figure 3.9. Effect of Mfd and an Mfd mutant defective in RNAP displacement on roadblock repression of pRCBlacO$_{ID}$luc.

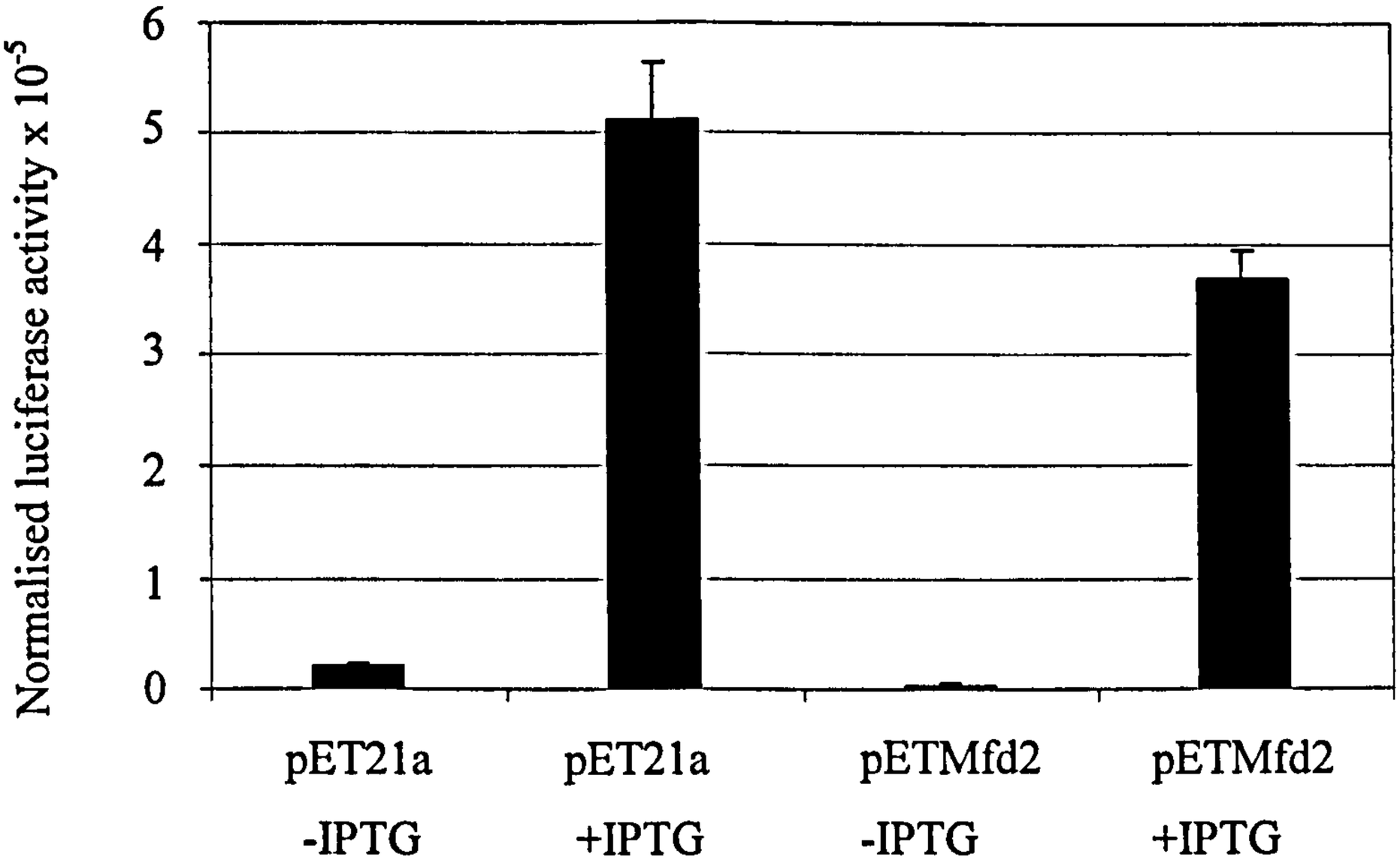
UNCNOMFD (*mfd*) cells containing the reporter construct pRCBlacO$_{ID}$luc were transformed with pET21a, pETMfd or pETMfd2Trunc$_{1-939}$. Cultures were grown to mid-exponential phase (OD_{600} ~0.5) in M9 media containing the appropriate antibiotics. Luciferase activity was measured and the normalised luciferase activity calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested. The values shown are an average of three independent experiments and are shown with s.d.

experiments were carried out to test repression of the reporter gene by Lac repressor and the effect of Mfd on this repression.

The effect of Lac repressor on repression of the luciferase reporter gene of pRCBKA4 was examined by transforming UNCNOMFD cells with pRCBKA4 and the control plasmid pET21a. Cultures were grown in M9 in the absence (repressed state) or presence (induced state) of IPTG to mid-exponential phase and the luciferase activity was measured (figure 3.10). In the repressed state, binding of Lac repressor to the two operator sites repressed transcription of the luciferase gene 25-fold in comparison to the induced state. This was approximately 2-fold tighter repression than was observed for the single operator reporter, pRCBlacO_{ID}luc and approximately 2.5-fold tighter repression than observed for pRCB-CAT. Therefore expression of the luciferase gene of pRCBKA4 is repressed by the binding of Lac repressor to operator sites downstream of the promoter.

To test whether Mfd increased the efficiency of roadblock repression of the luciferase gene in pRCBKA4, UNCNOMFD cells containing pRCBKA4 were transformed with pETMfd2. Cultures were grown to mid-exponential phase under repressing and inducing conditions and the luciferase activity was assayed (figure 3.10). The luciferase activity in the repressed state was reduced 123-fold in comparison to the induced state in cells containing Mfd, equivalent to a 4.8-fold Mfd effect on the repressed level of *luc* transcription.

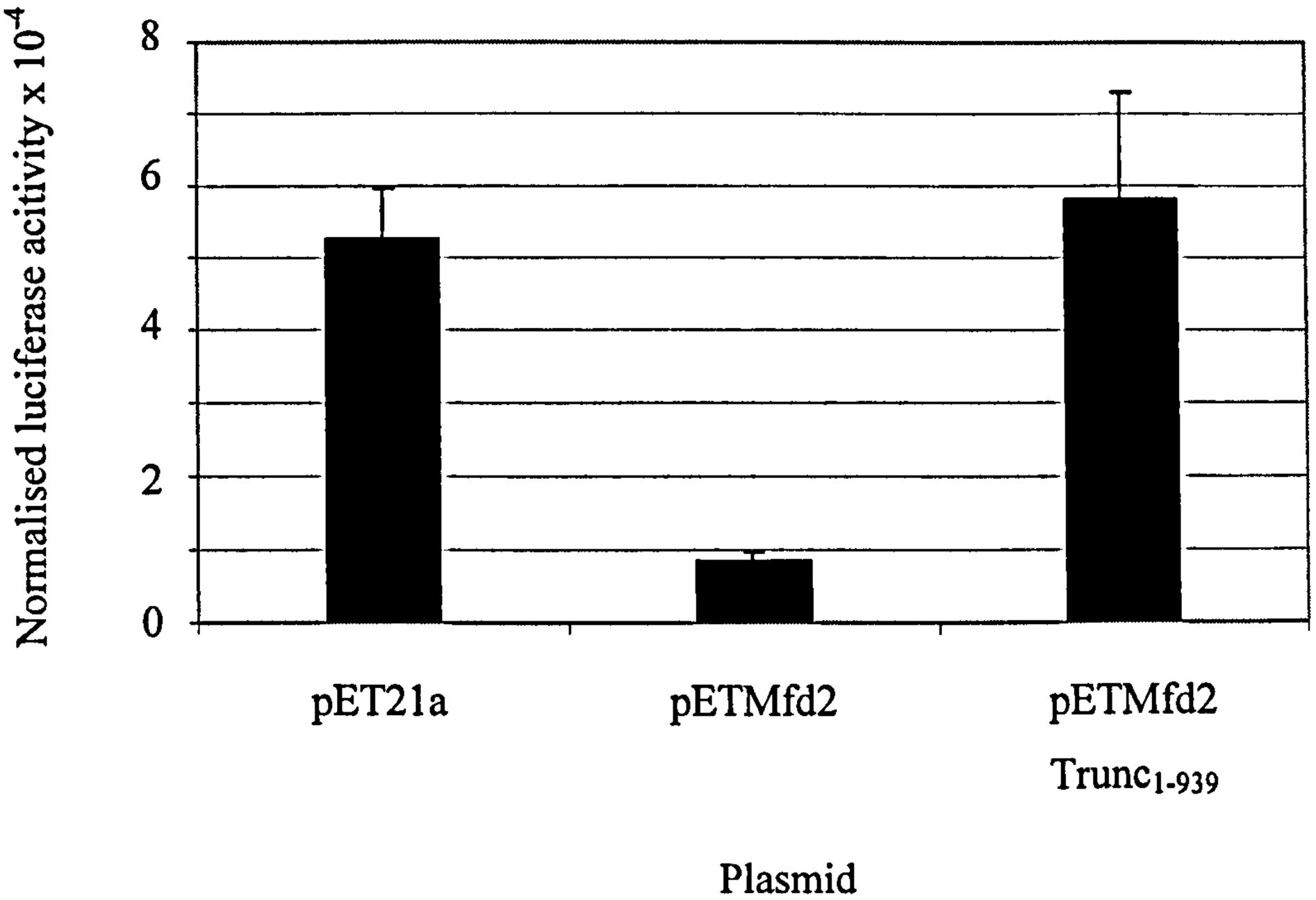
In order to examine if, as for pRCB-CAT, the increase in the efficiency of roadblock repression in the presence of Mfd was dependent on its ability to displace RNAP, UNCNOMFD cells containing pRCBKA4 were transformed with pET21a, pETMfd2 or pETMfd2Trunc$_{1\text{-}939}$. Cells were grown to mid-exponential phase and the luciferase activity was determined (figure 3.11). In the repressed state, Mfd



Mfd	+IPTG (induced)	-IPTG (repressed)	Repression (+IPTG/-IPTG)
None	5.1 ± 0.5	0.2± 0.05	25.5-fold
Wild-type	3.7 ± 0.2	0.03± 0.005	123-fold

Figure 3.10. Repression of pRCBKA4 by lac repressor.

UNCNOMFD (*mfd*) cells containing the reporter construct pRCBKA4 were transformed with either pETMfd or pET21a. Cultures were grown to mid-exponential phase (OD_{600}~0.5) in M9 media containing the appropriate antibiotics and where indicated 0.5 mM IPTG. Luciferase activity was measured and the normalised luciferase activity calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested. The values shown are an average of three independent experiments and are shown with s.d.



Expression Plasmid	Repressed normlised luciferase activity x10^{-4}	Mfd effect (-Mfd/+Mfd) in repressed state
pETMfd2	0.9±0.1	5.8-fold
pETMfd2Trunc$_{1-939}$	6.3±1.5	0.8-fold

Figure 3.11. Effect of Mfd and an Mfd mutant defective in RNAP displacement on roadblock repression of pRCBKA4 in *mfd*⁻ cells.

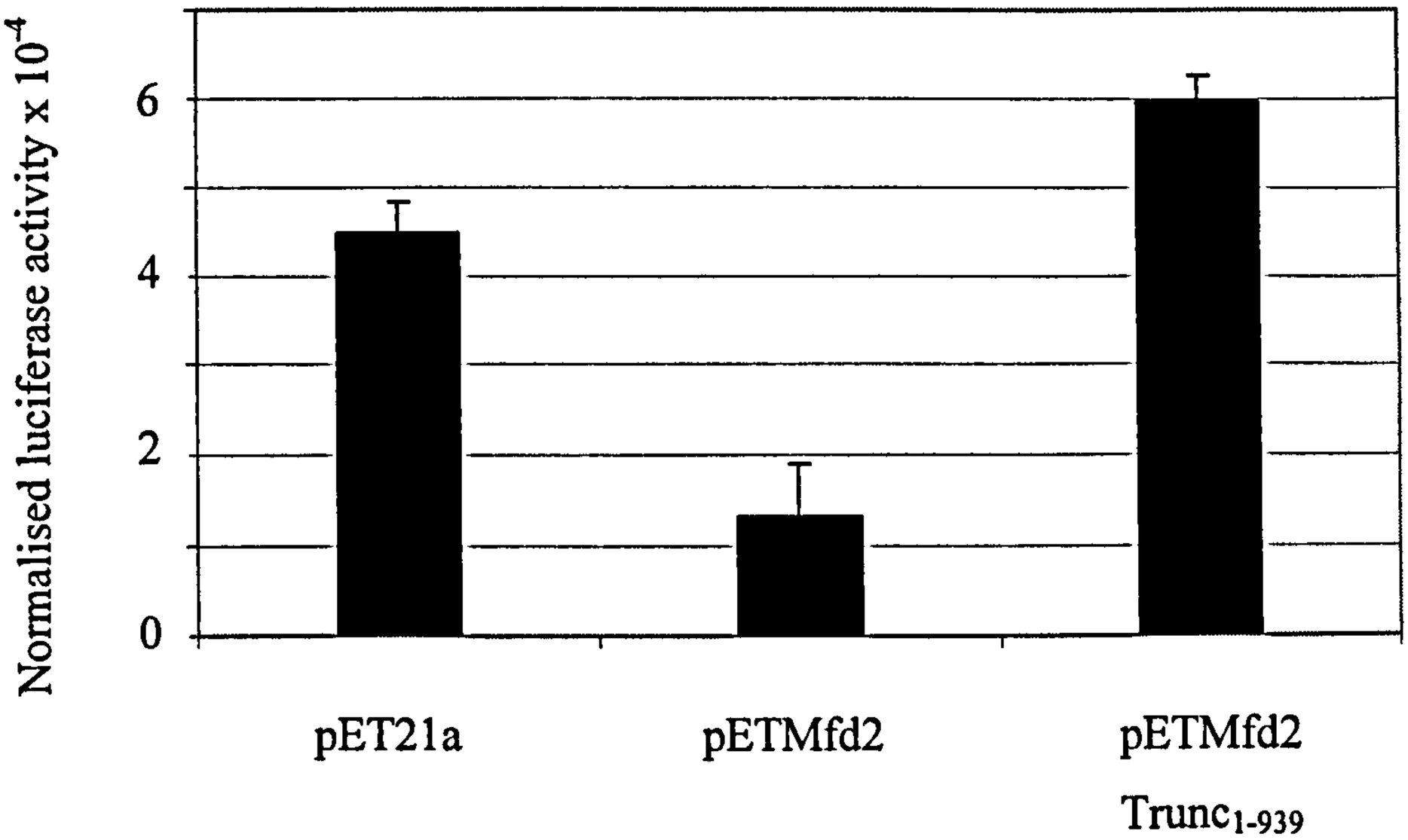
UNCNOMFD (*mfd*⁻) cells containing the reporter construct pRCBKA4 were transformed with pETMfd2, pETMfd2Trunc$_{1-939}$ or pET21a. Cultures were grown to mid-exponential phase (OD_{600}~0.5) in M9 media containing the appropriate antibiotics. Luciferase activity was measured and the normalised luciferase activity calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested. The values shown are an average of three independent experiments and are shown with s.d.

expression resulted in luciferase activity 5.8-fold lower than in cells lacking Mfd. Luciferase activity in cells containing pETMfdTrunc$_{1\text{-}939}$ was comparable to that of cells lacking the Mfd protein.

The results show that the ability of Mfd to increase the efficiency of roadblock repression of the luciferase gene in pRCBKA4 is dependent on being able to displace RNAP. These results are consistent with those obtained using the pRCB-CAT reporter system and indicate that pRCBKA4 is a suitable reporter construct to assay Mfd activity *in vivo*.

Using the CAT roadblock reporter system, MfdTrunc$_{1\text{-}939}$ showed a transdominant effect over endogenous Mfd. To test whether this was the case for the pRCBKA4 luciferase reporter system, AB1157 cells containing the roadblock reporter construct were transformed with plasmids encoding wild-type Mfd, MfdTrunc$_{1\text{-}939}$ or a plasmid without the *mfd* gene. The luciferase activity of cultures in mid-logarithmic phase was determined (figure 3.12). As observed for the CAT roadblock reporter system, over-expression of Mfd caused a reduction (3.5-fold) in reporter activity under repressing conditions compared to cells in which synthesis of Mfd only occurred from the chromosomal gene. Over-expression of MfdTrunc$_{1\text{-}939}$ caused a modest, 1.3-fold, increase in luciferase activity, confirming the transdominant effect observed in CAT assays.

Taken together these results indicate that the pRCBKA4 luciferase reporter is a suitable construct to assay Mfd function *in vivo*. The results obtained using the pRCBKA4 *luc* reporter system were comparable to those obtained using the pRCB-CAT *cat* reporter system. In both of these constructs Lac repressor binding to an operator site located between the promoter and the reporter gene repressed



Mfd Expression Plasmid	Repressed normalised luciferase activity x10^{-4}	Mfd effect (-Mfd plasmid/+Mfd plasmid) in repressed state
pETMfd2	1.3±0.6	3.5-fold
pETMfd2Trunc$_{1-939}$	5.9±0.3	0.8-fold

Figure 3.12. Effect of Mfd and an Mfd mutant defective in RNAP displacement on roadblock repression of pRCBKA4 in *mfd*$^+$ cells.

AB1157 (*mfd*$^+$) cells containing the roadblock repression reporter pRCBKA4 were transformed with the plasmid indicated. pET21a was used as a control plasmid to examine endogenous Mfd and pETMfd2 was used to over-express Mfd protein. Cultures were grown to an OD_{600} ~0.5 in M9 media containing the appropriate antibiotics. Luciferase activity was measured and the normalised luciferase activity calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested. The values shown are an average of three independent experiments and are shown with s.d.

expression of the reporter gene. Mfd increased the efficiency of roadblock repression of both of these reporter systems and this effect was dependent on the ability of the Mfd protein to displace stalled elongation complexes.

DISCUSSION

A roadblock repression reporter system was developed to examine whether Mfd contributed to roadblock repression in *E. coli,* to determine the mechanism of any such effect and to establish an assay for Mfd activity *in vivo*. Lac repressor bound to *lac* operator sites was used to form protein roadblocks in order to stall transcription upstream of a reporter gene. The efficiency of repression of transcription by this roadblock was greater in cells containing Mfd than in cells lacking Mfd. Mfd therefore increases roadblock repression *in vivo* in *E. coli* as observed previously for *B. subtilis* (Zalieckas *et al.*, 1998). Although the effect of Mfd on roadblock repression may be physiologically relevant for only a handful of operons in *E. coli,* it provides a means of assaying Mfd activity *in vivo*.

Mfd increased the efficiency of roadblock repression of two different reporter constructs containing tandem *lac* operators, 8-fold and 5.8-fold for *cat* and *luc* reporters respectively. The first of these constructs, pRCB-CAT, contained a *cat* reporter gene flanked by ideal *lac* operators. The fact that CAT activity is reduced in the presence of Mfd means that this system provides an assay for Mfd function *in vivo* and an easily monitored phenotype for *mfd*⁻ cells. Cells that do not contain functional Mfd grow readily on plates containing 5 μg/ml chloramphenicol whereas cells containing wild-type Mfd are not viable. The reporter activity in the luciferase roadblock repression system that was developed, using pRCBKA4, is also dependent on Mfd and therefore can be used to assay Mfd function *in vivo* from liquid culture. The effect of introducing an auxiliary operator only increased repression by Lac repressor ~2-fold in pRCBKA4 compared to the level of repression at the single *lac* operator in pRCBlacO_{ID}luc. Mfd had a greater effect on the efficiency of roadblock repression in cells containing the tandem operator construct than in cells containing

the reporter with a single operator. The presence of two operators may decrease dissociation of the Lac repressor and therefore provide a longer time in which Mfd can act on the stalled elongation complex before the repressor leaves the DNA and transcription continues. Alternatively, the two operators in pRCKA4 may each act as a separate stall site for RNA polymerase; therefore any elongation complexes that have not been displaced or stalled at the first roadblock may stall at the second operator site, providing a second chance for displacement by Mfd. This is not the case for the *cat* reporter construct pRCB-CAT where the second operator site is located at the end of the reporter gene. The induced level of reporter activity of both pRCB-CAT and pRCBKA4 is lower in the presence of Mfd than in its absence. This may be due to Mfd displacing elongation complexes stalled within the reporter gene at natural pause sites.

Previous work on the contribution of auxiliary operators and spacing between operator sites on repression by Lac repressor has been carried out in *mfd*$^+$ cells (Muller *et al.*, 1996). Therefore it is possible that some of the increased repression that is observed in these experiments upon the introduction of auxiliary operator sites is due to the contribution of Mfd.

In vitro work using Mfd_{trunc} had shown that this truncated protein was able to bind DNA and RNAP but was no longer able to displace stalled elongation complexes (Selby and Sancar, 1995a). $MfdTrunc_{1-939}$, truncated by a stop codon after amino acid 939, rather than by C-terminal fusion to the *lacZ* α-peptide used in Mfd_{trunc}, was unable to increase the efficiency of roadblock repair *in vivo*, presumably due to its inability to displace transcription complexes stalled at the roadblock. This suggests that the increased efficiency of roadblock repression in the presence of Mfd is dependent on its ability to displace RNAP. In addition, over-expression of

MfdTrunc$_{1\text{-}939}$ had a trans-dominant effect over endogenous Mfd, indicating that it competed with the wild-type protein for binding to stalled elongation complexes and therefore indicated that the truncated protein was expressed.

In summary, two *in vivo* assays for roadblock repression have been established using the *cat* and *luc* genes as reporters. The level of reporter gene expression in these systems is dependent on the Mfd status of the cell, with functional Mfd protein increasing the efficiency of roadblock repression. Moreover, by using a mutant that has been shown to be defective in stalled elongation complex displacement *in vitro,* the Mfd effect on reporter gene expression was found to require the RNAP displacement activity of Mfd. Therefore the increased efficiency of roadblock repression is likely to be caused by displacement of transcription complexes stalled at the Lac repressor roadblock located upstream of the reporter gene. By using the *cat* gene as a reporter the roadblock repression system provides a detectable phenotype for *mfd*$^-$ cells or cells containing non-functional Mfd. The chloramphenicol-resistant phenotype of cells containing this reporter construct but lacking the RNAP displacement activity of Mfd permits the isolation of mutants that are defective in this function of Mfd.

CHAPTER 4

THE TRG MOTIF OF Mfd

INTRODUCTION

A region of *E. coli* Mfd that encompasses the helicase motifs (amino acids 598-968) shares 38% identity with amino acids 266-645 of *E. coli* RecG. A recently discovered highly conserved motif, termed the TRG (Translocation in **RecG**) motif, also falls in this region of RecG. In the crystal structure of RecG (Singleton and Wigley, 2003) the TRG motif is shown to consist of two helices linked by a four amino acid loop to form a helical hairpin and a well-defined loop is located C-terminal to this hairpin. The whole of this region may be stabilised by a short helix that follows the loop, which crosses back over a long α helix that connects the N-terminal helicase domain to the junction-binding domain. Mutagenesis of the RecG TRG motif implicated the motif in coupling ATP binding and hydrolysis to translocation along double-stranded DNA (Mahdi *et al.*, 2003).

The region downstream of the helicase motifs in the Mfd protein shares sequence similarity with the TRG motif of RecG (figure 4.1). This region of Mfd is predicted to contain two α helices, although they are separated by a seven amino acid loop rather than the four amino acid loop present in RecG. The putative TRG motif of Mfd was truncated in $MfdTrunc_{1\text{-}939}$ and in order to determine whether the putative TRG motif of Mfd is required for Mfd function, a second C-terminal truncation of Mfd, retaining an intact putative TRG motif, was created. In addition, in order to identify residues in this region that are critical for Mfd function random mutagenesis and site-directed mutagenesis approaches were adopted. The mutants that were generated were assayed *in vivo* using the roadblock repression assay developed in chapter 3 and their DNA binding, ATP hydrolysis and elongation complex displacement activities were characterised *in vitro*.

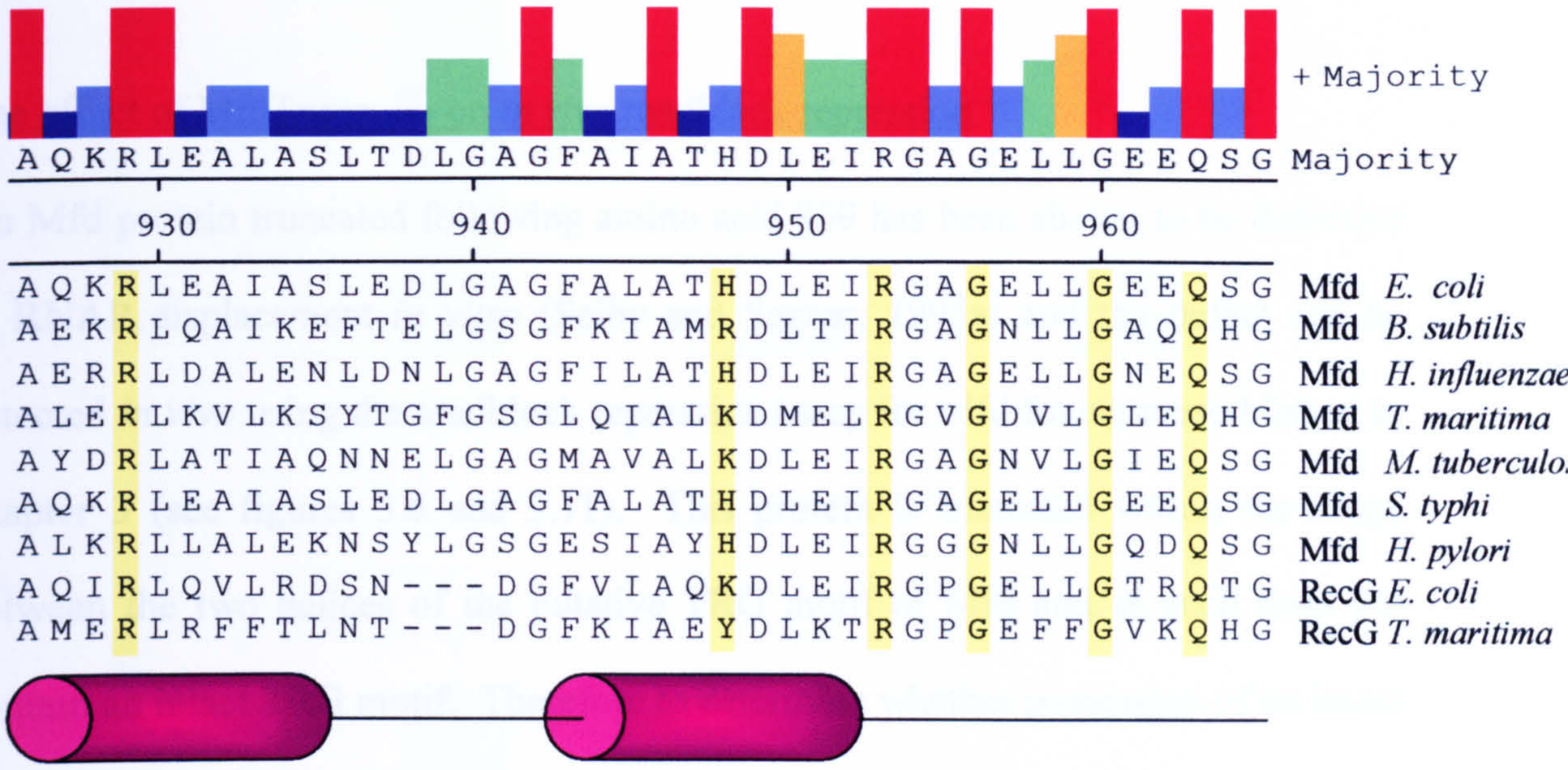


Figure 4.1. Sequence alignment of the putative TRG motif of Mfd

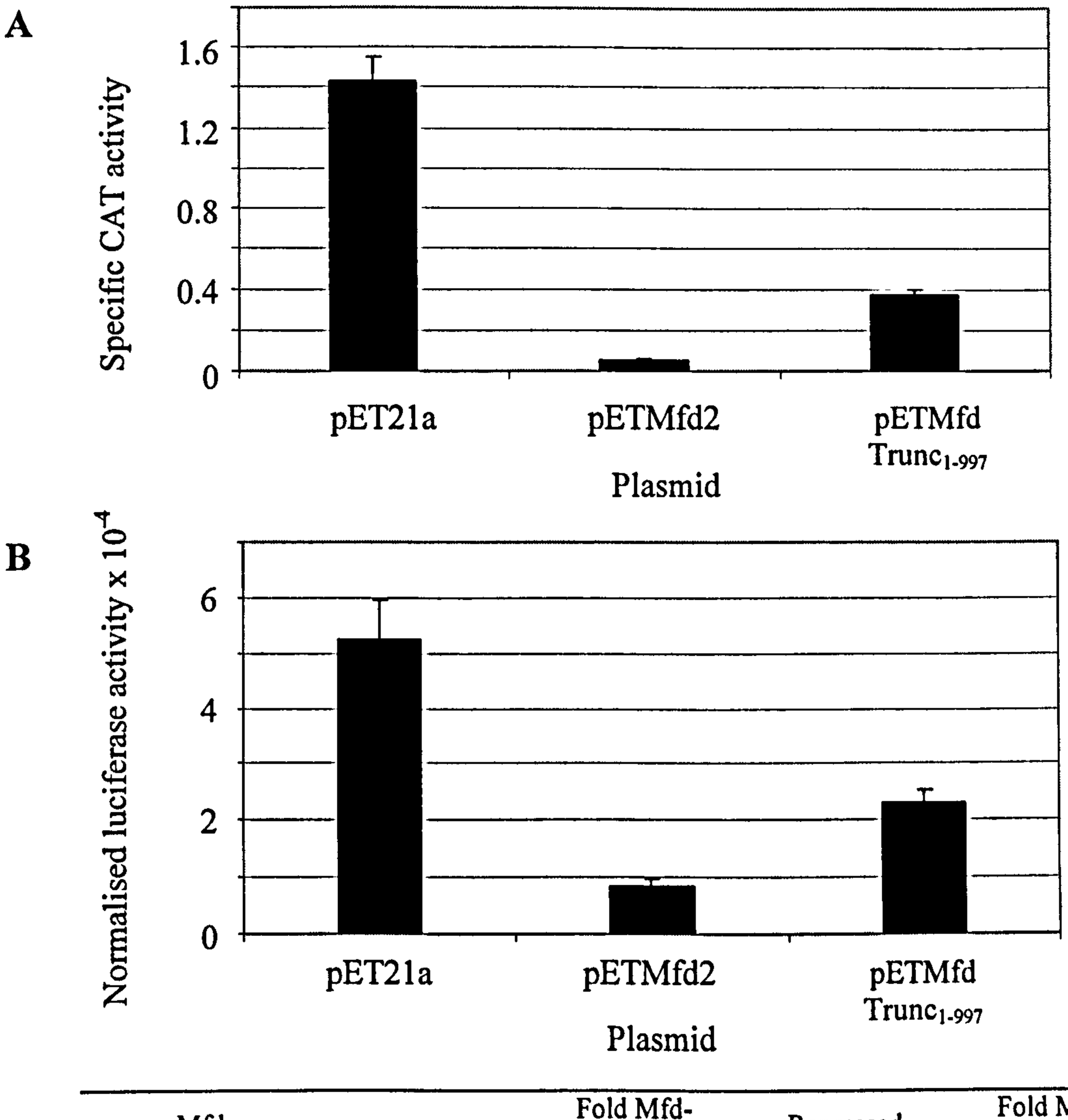
An alignment of a sample of Mfd protein sequences with RecG sequences from *E. coli* and *T. maritima*. The degree of conservation is indicated by the bar above, with a red bar indicating residues that are 100% identical. The residue numbering shown is for *E. coli* Mfd and the positions of the helices in the RecG *T. maritima* crystal structure are shown relative to the primary sequence. Residues substituted by random and site-directed mutagenesis in the experiments described in this thesis are highlighted in yellow.

RESULTS

The effect of MfdTrunc$_{1-997}$ on *in vivo* roadblock repression

An Mfd protein truncated following amino acid 939 has been shown to be defective in RNAP displacement *in vitro* (Selby and Sancar, 1995a) and the defect can be detected *in vivo* using the roadblock repression assay for Mfd function established in chapter 3 (see figures 3.6 and 3.11). This protein is truncated within the hinge between the two helices of the putative TRG motif of Mfd and as such does not contain an intact TRG motif. Therefore to determine whether possession of an intact putative TRG motif is sufficient to restore RNAP displacement activity, a construct expressing an Mfd protein with a shorter C-terminal deletion was constructed. Sequence alignment with RecG predicted that the TRG motif of Mfd is located between amino acids 926-965; therefore introduction of a stop codon following amino acid 997 of Mfd will result in a protein with an intact putative TRG motif. In the crystal structure of RecG the region immediately downstream of the TRG motif forms a short helix that crosses over a long helix, which connects the N-terminus of the helicase domains to the RecG junction-binding domain. The MfdTrunc$_{1-997}$ protein contains this helix, which may act to stabilise the helicase domains.

pETMfdTrunc$_{1-997}$, a derivative of pETMfd2, which expresses Mfd truncated after amino acid 997 was constructed and its ability to displace RNAP *in vivo* was tested in both the *cat* and the *luc* roadblock repression assays. UNCNOMFD (*mfd*$^-$) cells containing either the *cat* reporter construct pRCB-CAT (figure 4.2 A) or the *luc* reporter construct pRCBKA4 (figure 4.2 B) were transformed with the empty vector pET21a, pETMfd2 or pETMfdTrunc$_{1-997}$. Cultures were grown to mid-exponential phase and the activity of the reporter was determined (figure 4.2).



Mfd Expression Plasmid	Repressed CAT activity	Fold Mfd-effect in *cat* system (-Mfd/+Mfd)	Repressed luciferase activity	Fold Mfd-effect in *luc* system (-Mfd/+Mfd)
pETMfd2	0.05+/-0.01	27.8	0.86+/-0.12	6.14
pETMfdTrunc$_{1\text{-}997}$	0.37+/-0.03	3.87	2.32+/-0.20	2.26

Figure 4.2. The effect of Mfd truncated following amino acid 997 on roadblock repression *in vivo*.

UNCNOMFD (*mfd*) cells containing either the roadblock repression reporter pRCB-CAT (A) or pRCBKA4 (B) were transformed with pET21a, pETMfd2 or pETMfdTrunc$_{1\text{-}997}$. Cultures were grown in M9 media containing the appropriate antibiotics and cells were harvested at OD_{600} ~0.5. Values are the average of three independent experiments and are shown with s.d. Specific CAT activity units are nmol chloramphenicol acetylated/min/mg protein and normalised luciferase activity was calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested.

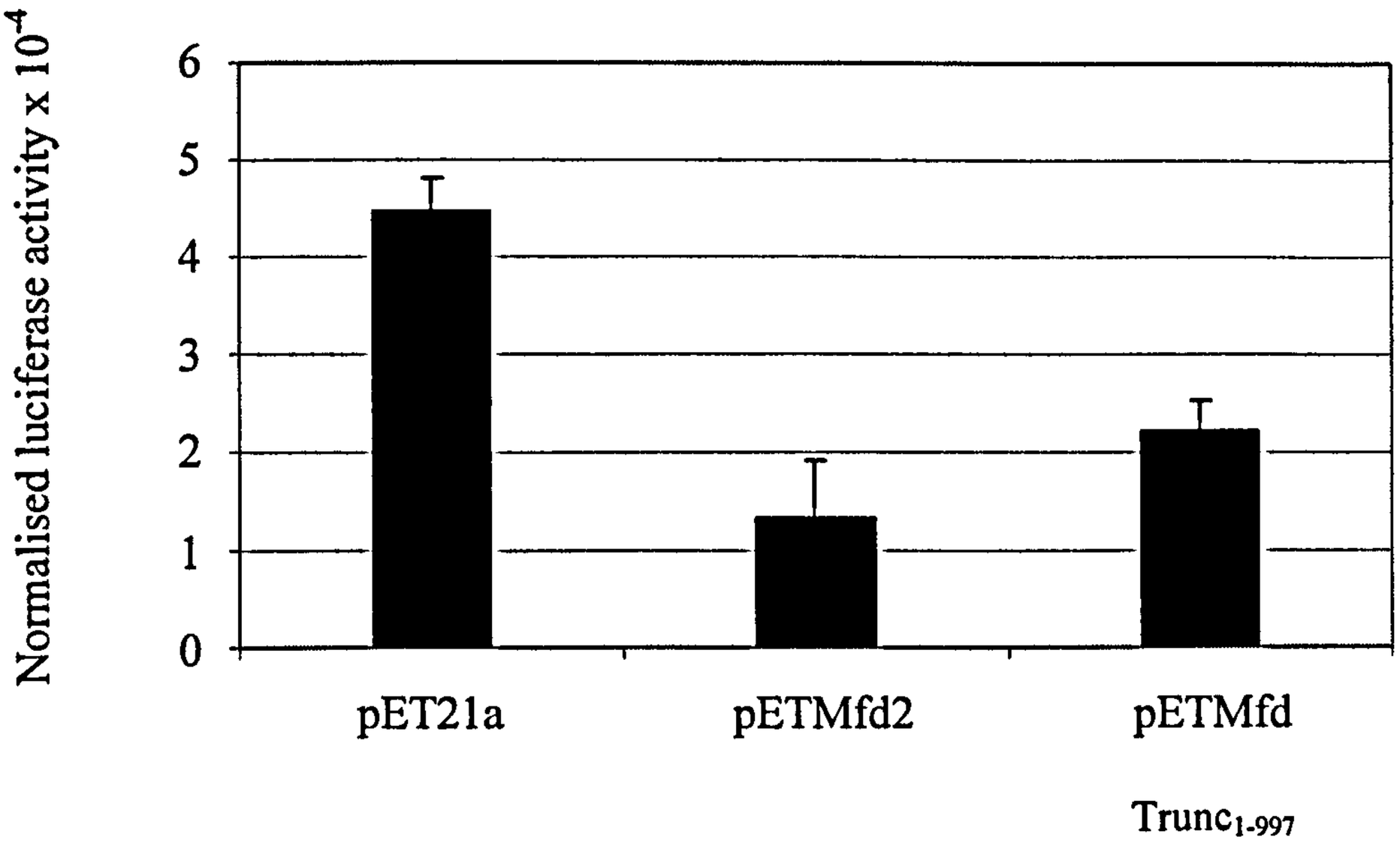
Cells containing pRCB-CAT and expressing $TruncMfd_{1\text{-}997}$ possessed reduced CAT activity in comparison to cells not expressing Mfd. The efficiency of roadblock repression was increased 3.9-fold in the *cat* reporter system whilst in cells containing the pRCBKA4 reporter and $pETMfdTrunc_{1\text{-}997}$ luciferase activity was 2.3-fold lower than in cells not expressing Mfd. In both reporter systems the reporter activity in the presence of $MfdTrunc_{1\text{-}997}$ was not reduced to quite the same level as in the presence of the wild-type protein.

These results showed that an Mfd protein possessing an intact TRG region was able to increase the efficiency of roadblock repression in both reporter systems, indicating that this protein is able to displace elongation complexes *in vivo*.

Mfd truncated following amino acid 939 exhibited a trans-dominant effect over endogenous Mfd, interfering with the ability of wild-type Mfd to displace elongation complexes. To examine whether truncation of Mfd following amino acid 997 and therefore whether Mfd protein containing a complete putative TRG motif shows a trans-dominant effect, assays were repeated in mfd^+ cells.

AB1157 (mfd^+) cells containing the *luc* roadblock reporter pRCBKA4 were transformed with pET21a, pETMfd2 and $pETMfdTrunc_{1\text{-}997}$. Cultures were grown to mid-exponential phase and the reporter activity was determined (figure 4.3). Over-expression of $MfdTrunc_{1\text{-}997}$ resulted in a 2-fold decrease in luciferase activity in comparison to cells containing only endogenous Mfd. $MfdTrunc_{1\text{-}997}$ increased the efficiency of roadblock repression efficiency in mfd^+ cells slightly less than wild-type Mfd, which caused a 3.3-fold decrease in luciferase activity.

Over-expression of $MfdTrunc_{1\text{-}997}$ had a slightly smaller effect on roadblock repression than over-expression of the wild-type protein in both mfd^+ and mfd^- cells. This modest defect could be due to a decreased rate of elongation complex



Mfd Expression Plasmid	Repressed Luciferase activity	Fold Mfd effect (-Mfd plasmid/+Mfd plasmid)
pETMfd2	1.34+/-0.58	3.35
pETMfdTrunc$_{1-997}$	2.23+/-0.26	2.01

Figure 4.3. The effect of Mfd truncated following amino acid 997 on roadblock repression *in vivo* in *mfd*$^+$ cells.

AB1157 (*mfd*$^+$) cells containing the roadblock repression reporter pRCBKA4 were transformed with pET21a, pETMfd2 or pETMfdTrunc$_{1-997}$. Cultures were grown in M9 media containing the appropriate antibiotics and cells were harvested at OD_{600} ~0.5. Values are the average of three independent experiments and are shown with s.d. Normalised luciferase activity was calculated by dividing the luciferase reading by the OD_{600} at which the cells were harvested.

displacement, decreased expression or stability of the protein in comparison to the wild-type or an effect on another aspect of Mfd function such as whether the Mfd protein remains bound to the DNA or RNAP following elongation complex displacement.

Together the results show that Mfd protein that is truncated after residue 997 is still able to displace elongation complexes *in vivo*. The putative TRG motif is intact in this protein but is disrupted in $MfdTrunc_{1\text{-}939}$, which failed to displace RNAP *in vivo*. This suggests that the 58 amino acid region of Mfd between 939 and 997 is required for dissociation of stalled transcription elongation complexes and therefore it seems likely that the putative TRG motif of Mfd is involved in elongation complex displacement.

Random mutagenesis of the putative TRG motif of Mfd

To further examine the role of the putative TRG motif of Mfd in elongation complex displacement and furthermore to determine which residues are essential for this function, the *in vivo* roadblock repression assay developed in the previous chapter can be used to perform a non-biased screen for mutants of *mfd* that are defective in displacement of elongation complexes stalled at a protein roadblock. In cells expressing wild-type Mfd, RNAP stalled at the Lac repressor block is displaced and therefore little transcription of the downstream *cat* reporter gene occurs. In cells expressing Mfd derivatives that are unable to displace RNAP, transcription can resume upon dissociation of the repressor protein and so the *cat* gene is expressed. As a result, clones expressing Mfd defective in elongation complex displacement are chloramphenicol-resistant (figure 3.4).

The screening strategy that was employed is shown in figure 4.4. Error-prone PCR was used to introduce random mutations into the region of the *mfd* gene that contains

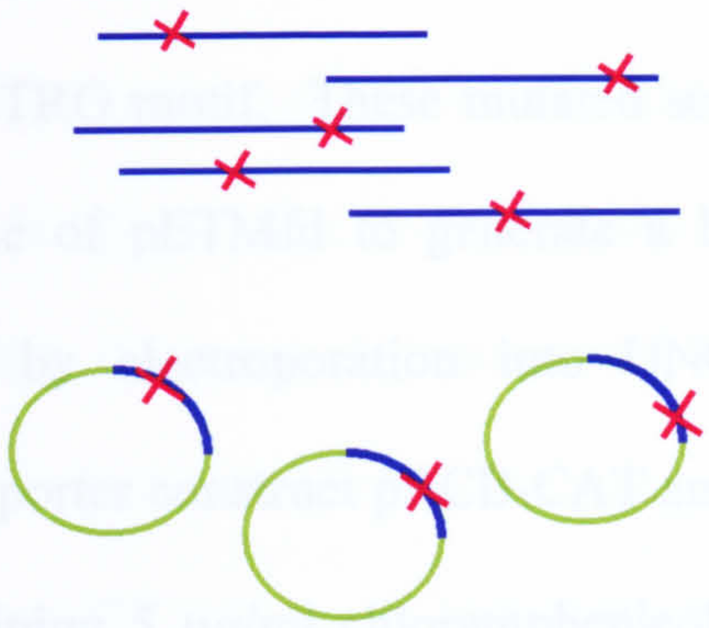


Figure 4.4. Random mutagenesis and screening of the TRG region of Mfd.

The screening strategy for identification of substitutions of Mfd that result in an inability to increase roadblock repression efficiency *in vivo* was divided into 4 stages. Stage 1 was the identification of chloramphenicol-resistant clones, stage 2 was the confirmation of this phenotype by restreaking, stage 3 was the isolation of *mfd* plasmid of the correct size and stage 4 was the confirmation that the chloramphenicol-resistant phenotype was conferred by the *mfd* plasmid, and sequencing of the plasmid.

the putative TRG motif. These mutated sections of the *mfd* gene were cloned into the *mfd* gene of pETMfd to generate a library of mutants. The libraries were transformed by electroporation into UNCNOMFD (*mfd*$^-$) cells containing the roadblock reporter construct pRCB-CAT and the transformed cells were spread onto plates containing 5 μg/ml chloramphenicol (stage 1). The ligation efficiency and frequency of mutations that resulted in chloramphenicol-resistance was established by also spreading the transformations onto plates that did not contain chloramphenicol. The colonies that grew on plates containing 5 μg/ml chloramphenicol were restreaked to confirm their antibiotic resistance (stage 2) and to ensure that plasmid DNA was extracted from a single colony. The plasmid preparations from the restreaked colonies contained the reporter plasmid in addition to the mutated *mfd* plasmid. DNA was run on an agarose gel to ensure the Mfd plasmid was of the correct size and had not undergone large-scale deletions or insertions that could have resulted in a chloramphenicol-resistant phenotype (stage 3). In order to remove the reporter plasmid, the preparation was incubated with *Nco*I, which cuts only the reporter plasmid, before being transformed into a) UNCNOMFD cells containing pRCB-CAT and b) XL1-Blue cells. The UNCNOMFD cells were spread onto agar containing 5 μg/ml chloramphenicol to confirm that the chloramphenicol-resistant phenotype was conferred by the mutant *mfd* plasmid while plasmid DNA was extracted from XL1-Blue cells so the sequence of the Mfd plasmid could be determined (stage 4).

Initial screening procedure

The first screening approach that was adopted (A) utilised a method of error-prone PCR that takes advantage of the natural error frequency of *Taq* polymerase (8×10^{-6} errors/bp) (Cline *et al.*, 1996). PCR was carried out using a protocol that contained

40 amplification cycles and that used pUC-Mfd, which encodes amino acids 895-997 of Mfd, as a template. The reverse primer used contained a *Bsm*AI site, which was positioned such that upon digestion with *Bsm*AI a *Bgl*II end was generated. This meant that the mutated region of the *mfd* gene encoding amino acids 913-997 could be cloned into pETMfd as an *Nsi*I-*Bgl*II fragment to generate libraries of mutated *mfd* genes.

Libraries generated in this manner from six independent PCR reactions were transformed by electroporation into UNCNOMFD cells containing pRCB-CAT and were spread onto LB agar plates containing 0.4% glucose, kanamycin, tetracycline, ampicillin and 5 μg/ml chloramphenicol. To determine the efficiency of ligation and to estimate the number of colonies that were screened, transformations were also spread onto LB agar plates containing 0.4% glucose, kanamycin, tetracycline, ampicillin but lacking chloramphenicol. From the number of colonies that grew in the absence of chloramphenicol it was calculated that approximately 1500 clones were screened for chloramphenicol resistance. The ligation efficiency of the cloning procedure was low, probably due to incomplete digestion of the PCR products by *Bsm*AI and difficulty in generating large quantities of partially-digested *Nsi*I-*Bgl*II vector. Of the ~1500 clones screened on plates containing 5 μg/ml chloramphenicol, 70 showed a chloramphenicol-resistant phenotype (table 4.1, stage 1). This equated to 4.7% of the clones screened being chloramphenicol-resistant. The mutation frequency would have actually been higher than this since some of the mutations would have been silent or resulted in substitution of an amino acid that is not essential for RNAP displacement. Of the chloramphenicol-resistant colonies, 62 grew well when restreaked onto LB plate containing 0.4% glucose, kanamycin, tetracycline, ampicillin and chloramphenicol (table 4.1, stage 2). When the plasmid

Screening Stage		Template: pUC-Mfd (A)	Template: pBSMfd-TRG (B)	Total
	Number of colonies screened	~1500	~7500	~9000
1	Number of chloramphenicol-resistant colonies	70	179	249
2	Number of chloramphenicol-resistant clones following restreak	62	155	217
3	Number of plasmids of correct size	13	27	40
4	Number of sequences determined	4	17	21
	Single amino acid substitutions obtained	HL948	QL898 GD956 GV960	4 single amino acid substitutions

Table 4.1. Random mutagenesis screening of the TRG motif of Mfd.

The number of clones that reached each stage of the screening protocol in figure 4.4 is shown in the table above. Approximately 9000 clones were screened in total. Two different mutagenesis protocols were utilised. (A) pUC-Mfd containing the region of Mfd that encodes amino acids 895-997 was used as a template for error-prone PCR by amplification with Taq for 40 cycles. The resulting PCR product was digested with *Nsi*I and *Bsm*AI to leave a fragment that encoded amino acids 913-997. This was cloned into a partially digested *Nsi*I-*Bgl*II pETMfd vector. The library was screened on LB + 0.4% glucose agar plates containing 5 μg/ml chloramphenicol plus kanamycin, tetracycline and ampicillin. (B) Mutagenesis was carried out using a PCR method that involves unequal dNTP concentrations, Mn^{2+} and either 5, 7, 10 or 20 mutagenic PCR cycles. pBSMfd-TRG, which contains amino acids 850-997 of Mfd, was used as the PCR template. The product was cloned as a *Bsr*GI-*Bgl*II fragment, encoding amino acids 868-995, into pETMfd. The library transformations were plated onto M9 agar plates containing 5 μg/ml chloramphenicol plus kanamycin, tetracycline and ampicillin.

DNA was extracted from these colonies only 13 contained a pETMfd plasmid of the correct size (table 4.1, stage 3). The remainder possessed pETMfd plasmids that contained large deletions or insertions. The sequence of the mutated region of *mfd* from 4 of the plasmids that appeared to be the correct size was determined (table 4.1, stage 4) and of these one possessed a single amino acid substitution, HL948.

An alignment of 228 Mfd protein sequences obtained from the Entrez database showed that histidine is conserved in this position in 47% of the sequences. In 93% of the sequences the amino acid at this position is histidine, arginine or lysine i.e. is potentially positively charged. The other three plasmids that were sequenced had no mutations within the region that had been subject to mutagenesis and presumably these plasmids had acquired mutations elsewhere in the pETMfd plasmid, either within the Lac repressor gene or in a different region of the *mfd* gene.

Second screening procedure

The ligation efficiency of the first screening procedure was low and required transformation of many ligations in order to screen a reasonable number of clones. It was suspected that the poor ligation efficiency was due to the difficulty in producing large quantities of the partially-digested *Nsi*I-*Bgl*II pETMfd vector. Therefore a second method of generating mutant *mfd* libraries was adopted (table 4.1 B).

The region of *mfd* encoding amino acids 850-997 was cloned into pBluescript to generate the plasmid pBSMfd-TRG. This plasmid was used as the template for error-prone PCR using a protocol that included Mn^{2+} and unequal dNTP concentrations (Greene J, 2002). *Taq* polymerase was used for 20, 10, 7 or 5 amplification cycles of error-prone PCR, followed by cycles of high-fidelity PCR using Pfu polymerase to amplify the DNA without introducing any further mutations. The fragment encoding amino acids 868-995 was cloned as a *Bsr*GI-

*Bam*HI fragment into the *Bsr*GI and *Bam*HI sites of pETMfd. In addition to the putative TRG motif, this region of *mfd* includes part of helicase motif V and helicase motif VI. Libraries from 18 independent PCR reactions were screened in UNCNOMFD cells containing pRCB-CAT and the transformations were spread on M9 plates containing kanamycin, tetracycline, ampicillin and 5 μg/ml chloramphenicol. Approximately 7500 clones were screened in this manner and 179 chloramphenicol-resistant colonies were obtained, equivalent to 2.4% of the colonies screened (table 4.1, stage 1). Of these, 155 colonies grew well when restreaked onto chloramphenicol plates (table 4.1, stage 2). Only 27 of these colonies contained *mfd* plasmids that were the correct size and did not contain large-scale deletions or insertions (table 4.1, stage 3). The sequence of 17 of these plasmids was determined (table 4.1, stage 4) and 3 single amino acid substitutions were identified, QL898, GD956 and GV960, all of which were isolated from a screen using 5 cycles of error-prone PCR. G956 and G960 are located within the loop that follows the helical hairpin of the putative TRG motif while Q898 is located within helicase motif VI. An alignment of 228 Mfd sequences showed that Q898, G956 and G960 are all 100% conserved. The sequences obtained from the remaining 17 clones contained either no mutations within the region of *mfd* that was screened, frameshifts or multiple substitutions (table 4.2). From libraries generated by 20 cycles of error-prone PCR, five clones were obtained with multiple amino acid substitutions; these clones possessed an average of 5 substitutions within the 127 amino acid region that was subject to mutagenesis. The remainder of the plasmids sequenced, from screens using 5, 7 or 10 cycles of error-prone PCR, contained frameshift mutations or no mutations in the region of *mfd* that had been subject to mutagenesis. The predominant base change was substitution of A with T and no A to C, G to C, T to

A

No. of substitutions	Codon changes			
Frameshift (4 mutants)	**G942** GGT→GG- (fs)			
	G874 GGG→-GG (fs)	**PL919** CCA→CTA		
	K928 AAA→-AA (fs)	**TN969** ACC→AAC	**E991** GAG→GAA (s)	**L994** CTG→TTG (s)
	I884 ATT→-TT (fs)	**IT870** ATT→ACT	**EV872** GAA→GTA	**KE920** AAA→GAA
	EG958 GAA→GGA	**ED961** GAA→GAT		
Single (4 mutants)	**HL948** CAC→CTC			
	QL898 CAG→CTG	**L979** CTG→CTC (s)		
	GD956 GGT→GAT			
	GV960 GGC→GTC			
3 (1 mutant)	**IF871** ATC→TTC	**LQ949** CTG→CAG	**EV958** GAA→GTA	
4 (2 mutants)	**DV876** GAC→CTC	**LH959** CTT→CAT	**ED961** GAA→GAT	**IF970** ATC→TTC
	TA882 ACT→GCT	**VD903** GTC→GAC	**HQ907** CAT→CAA	**IV970** ATC→GTC
	R887 CGC→CGT (s)	**F891** TTC→TTT (s)		
6 (1 mutant)	**NI881** AAC→ATC	**IT885** AAT→ACT	**KN920** AAA→AAT	**KE928** AAA→GAA
	ED957 GAA→GAT	**SP973** TCG→CCG	**L915** TTG→CTG (s)	**T879** ACA→ACT (s)
7 (1 mutant)	**TI869** ACC→ATC	**IV885** ATT→GTT	**LQ893** CTG→CAG	**LP899** CTG→CCG
	TA916 ACA→GCA	**LP950** CTG→CCG	**ED961** GAA→GAT	**V983** GTC→GTT (s)
	D984 GAT→GAC (s)			

Eight of the mutants contained no mutations within the region of *mfd* that was sequenced.

B

Base change	Frameshift (single bp deletion)	A→T	A→G	T→A	T→C	C→A	C→T	G→A	G→T
No. of changes	4	14	6	5	7	1	6	3	1
% of changes	9	28	13	11	15	2	13	6	2

Table 4.2. Mutations obtained from screening of the putative TRG motif of Mfd.

(A) Table showing the mutants obtained from random mutagenesis screening of the putative TRG motif of Mfd. Each line indicates a separate clone. Single base deletions causing frameshifts are shown as (fs) while silent base changes are shown as (s).

(B) The comparative frequency of each base change in the mutants that were obtained. No A to C, G to C, C to G or T to G mutations were observed.

G or C to G mutations were obtained. This suggests that not all base changes are equally likely with the error-prone PCR protocol that was used.

Overall four single amino acid substitutions within the putative TRG motif of Mfd that resulted in defects in roadblock repression *in vivo* were isolated from libraries of randomly generated mutants. Approximately 9000 clones were screened to identify 249 chloramphenicol-resistant colonies. The screen of the putative TRG motif was non-saturating since none of the mutants was isolated twice. Relatively few single amino acid substitutions were isolated and this may be due to the relative infrequency of changes that abolish the function of an amino acid essential for RNAP displacement. A high background level of other mutations that generate the same phenotype may have masked these substitutions. Since the screen identifies loss-of-function substitutions that affect elongation complex displacement, a chloramphenicol-resistant phenotype will result from mutations elsewhere in the *mfd* gene that affect displacement, promoter mutations that decrease expression levels, substitutions that result in decreased stability of the Mfd protein or mutations in $lacI^q$ that affect Lac repressor expression or binding to DNA. Many of the *mfd* plasmids that were isolated contained large-scale deletions and this may well have lead to the loss of functional Lac repressor protein resulting in a chloramphenicol-resistant phenotype.

<u>Site-directed mutagenesis of the putative TRG motif of Mfd</u>

Site-directed mutagenesis can be used to examine whether the residues required for a functional TRG motif in RecG are also essential for TRG function in Mfd and therefore site-directed mutagenesis of the putative TRG motif of Mfd was carried out alongside the random mutagenesis described above. Residues were selected for substitution based on mutagenesis studies of the TRG motif of RecG (Mahdi *et al.*,

2003). Mutation of either of two juxtaposed arginine residues (R609 or R630) (figure 4.1) within the helical hairpin of the RecG TRG motif resulted in defects in activity both *in vivo* and *in vitro* (see introduction). *In vivo* the mutants were defective in complementing the UV sensitivity of a *recG* strain, while *in vitro* they were deficient in DNA unwinding. Substitution of R609 had more modest effects than substitution of R630 but acted synergistically with R630 in a double mutant to result in a protein with more severe defects than the single R630 mutant. Substitution of a glutamine residue (Q640) within the loop following the helical hairpin was shown to have an effect *in vivo* but its properties were not examined *in vitro*. The equivalent residues in *E. coli* Mfd are R929, R953 and Q963. R953 and Q963 are 100% conserved in all of the 228 Mfd sequences analysed while R929 is conserved in all but one of these sequences (*R. rickettsii*), where it is replaced by a cysteine residue.

Derivatives of pETMfd encoding alanine substitutions of residues R929, R953 and Q963 were created and the ability of the resulting Mfd proteins to dissociate stalled elongation complexes was tested *in vivo*. The pETMfd derivatives generated by site-directed mutagenesis, along with pETMfd expressing wild-type *mfd* and the empty vector pET21a, were transformed into UNCNOMFD cells containing the roadblock reporter construct pRCB-CAT. The transformations were spread on M9 plates containing 5 μg/ml chloramphenicol to determine whether the substitutions resulted in chloramphenicol-resistance in the roadblock repression assay. Cells containing the empty vector, and therefore no Mfd protein, grew readily on the plates containing chloramphenicol whilst the cells containing pETMfd expressing wild-type Mfd protein were chloramphenicol-sensitive. Cells containing Mfd RA929, Mfd RA953 and Mfd QA963 grew on plates containing 5 μg/ml

chloramphenicol, indicating that they were defective in displacement of an elongation complex stalled at a protein roadblock *in vivo* (figure 4.5).

Effect of substitutions in the TRG motif on Mfd activity in liquid culture

The effect of the mutants generated by random mutagenesis and site-directed mutagenesis on *in vivo* roadblock repression was tested quantitatively in liquid culture. UNCNOMFD (*mfd*) cells were transformed with the roadblock reporter construct pRCB-CAT and either pET21a or pETMfd plasmids encoding the wild-type Mfd protein or Mfd proteins containing QL898, RA929, HL948, RA953, GD956, GV960 or QA963 substitutions. The CAT activity in the repressed state of cultures grown to mid-exponential phase was determined (figure 4.5). Extracts from cells containing pETMfd derivatives encoding QL898, RA953, GD956, GV960 and QA963 substitutions possessed a level of CAT activity equivalent to that of cells containing no Mfd protein. These data show that the Mfd proteins containing these substitutions failed to increase the efficiency of roadblock repression and therefore suggest that Q898, R953, G956, G960 and Q963 are important residues in a function of Mfd that is required for Mfd-mediated elongation complex displacement. Subsequent purification of Mfd RA953 and Mfd QA963 from UNCNOMFD cells (see section on the effect of mutations in the TRG motif on activity *in vitro*) with a similar yield to the wild-type protein, confirmed that the absence of Mfd displacement activity is unlikely to be due to decreased stability or expression.

The CAT activity of cells containing pETMfdRA929 or pETMfdHL948 was equivalent to that of cells containing the wild-type pETMfd plasmid. These substitutions do not appear to affect the activity of Mfd in the roadblock assay in liquid culture and therefore these Mfd proteins cannot have completely lost RNAP

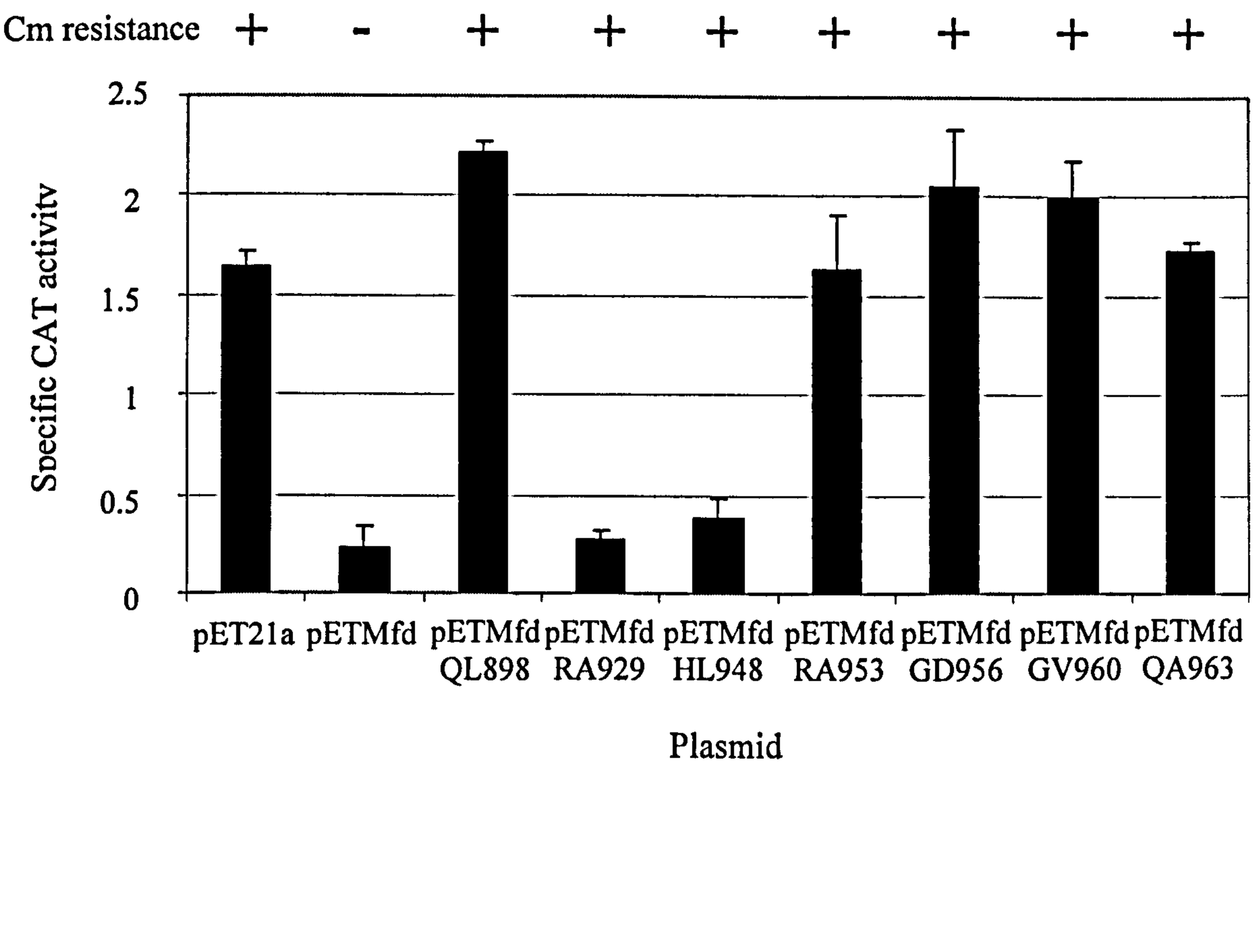


Figure 4.5. The *in vivo* effect of single amino acid substitutions within the putative TRG motif of Mfd on roadblock repression in *mfd*⁻ cells.

UNCNOMFD cells containing the roadblock repression reporter pRCB-CAT were transformed with the plasmid indicated. pET21a was used as a control plasmid that did not express Mfd. Cultures were grown in M9 media containing the appropriate antibiotics. Cells were harvested at OD_{600} ~0.5 and CAT activity determined. Values are the average of three independent experiments and are shown with s.d. Specific CAT activity units are nmol chloramphenicol acetylated/min/mg protein. The ability of transformed cells to grow on M9 agar plates containing ampicillin, kanamycin, tetracycline and 5 μg/ml chloramphenicol is indicated.

displacement activity. Whilst both RA929 and HL948 substitutions in Mfd conferred a chloramphenicol-resistant plate phenotype to *mfd*⁻ cells containing pRCB-CAT, extracts from liquid culture possessed the same CAT activity as cells containing wild-type *mfd*. The inconsistency may be due to a difference in sensitivity between the two assays or due to a difference in physiology between cells on solid media and in liquid culture.

If the mutant Mfd proteins are still able to bind to stalled RNAP but are defective in displacing it, they may impede the activity of endogenous Mfd. The ability of the substitutions to exert a trans-dominant effect over endogenous Mfd protein was tested *in vivo* using the roadblock repression system. AB1157 cells containing an intact chromosomal copy of the *mfd* gene were transformed with both the roadblock reporter plasmid pRCB-CAT and either the empty vector pET21a, pETMfd over-expressing wild-type Mfd or pETMfd derivatives encoding RA929, HL948, RA953 or QA963 substitutions. These substitutions were selected as being representative of the two classes of mutant that were observed in roadblock repression assays in *mfd*⁻ cells. RA929 and HL948 acted as wild type in liquid culture but exhibited chloramphenicol-resistance on agar plates whereas RA953 and QA963 were defective in both assays. The CAT activity of mid-exponential phase cultures was assayed (figure 4.6). As observed in chapter 3 (figure 3.6) the over-expression of Mfd from pETMfd resulted in lower CAT activity than when Mfd was only expressed from the chromosome. Mfd RA953 conferred a small but reproducible increase in CAT activity in AB1157 cells compared to cells transformed with pET21a, suggesting that like $MfdTrunc_{1\text{-}939}$, the protein is able to compete for binding sites on RNAP but fails to displace it. Mfd QA963, Mfd

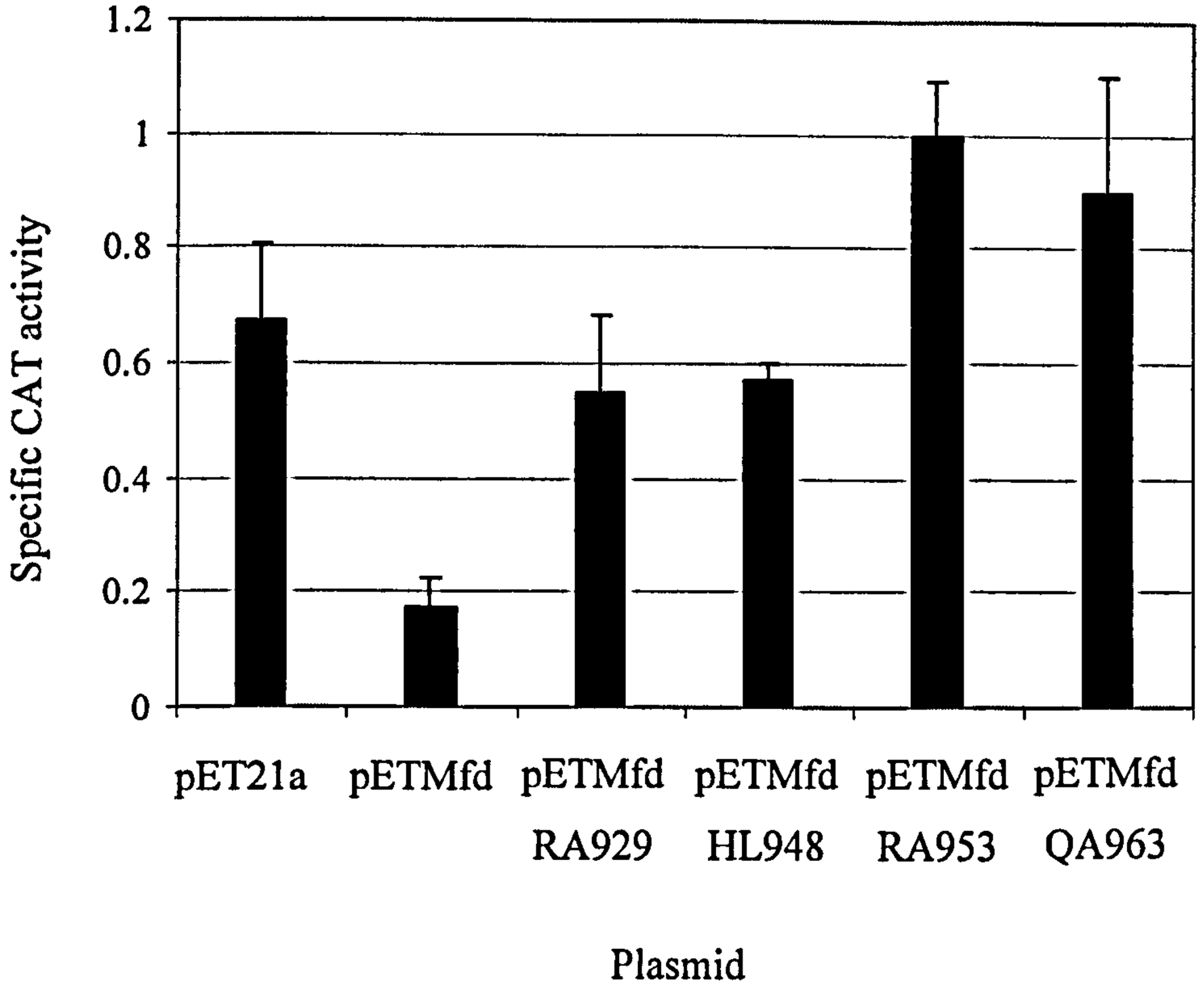


Figure 4.6. The *in vivo* effect of single amino acid substitutions within the putative TRG motif of Mfd on roadblock repression in *mfd*$^+$ cells.

AB1157 cells containing the roadblock repression reporter pRCB-CAT were transformed with the plasmid indicated. pET21a was used as a control plasmid not expressing Mfd. Cultures were grown in M9 media containing the appropriate antibiotics. Cells were harvested at OD_{600} ~0.5 and CAT assays were performed on triplicate cultures. The average CAT activity is shown with s.d.

RA929 and Mfd HL948 did not produce a statistically significant increase in CAT expression compared to the CAT expression in cells expressing Mfd only from the chromosome.

The effect of substitutions in the putative TRG motif of Mfd *in vitro*

Investigation of the *in vitro* properties of Mfd proteins containing substitutions within the TRG motif that affected the ability to displace elongation complexes *in vivo* will provide a greater understanding of the mechanism of RNAP displacement and the role of the TRG motif in this process. Mfd proteins that are unable to increase the efficiency of roadblock repression *in vivo* could have a defect in one of several activities. The proteins may be unable to bind to DNA, have lost the ability to bind or hydrolyse ATP or may be specifically defective in elongation complex displacement or coupling ATP hydrolysis to displacement. To determine which of these properties were affected by substitutions in the TRG motif that inhibited RNAP displacement *in vivo*, the ability of the mutant proteins to hydrolyse ATP, bind DNA and displace elongation complexes was tested *in vitro*. The Mfd RA929, HL948, RA953 and QA963 proteins were purified from UNCNOMFD cells by the method described in chapter 2.

ATPase activity of Mfd proteins containing TRG substitutions

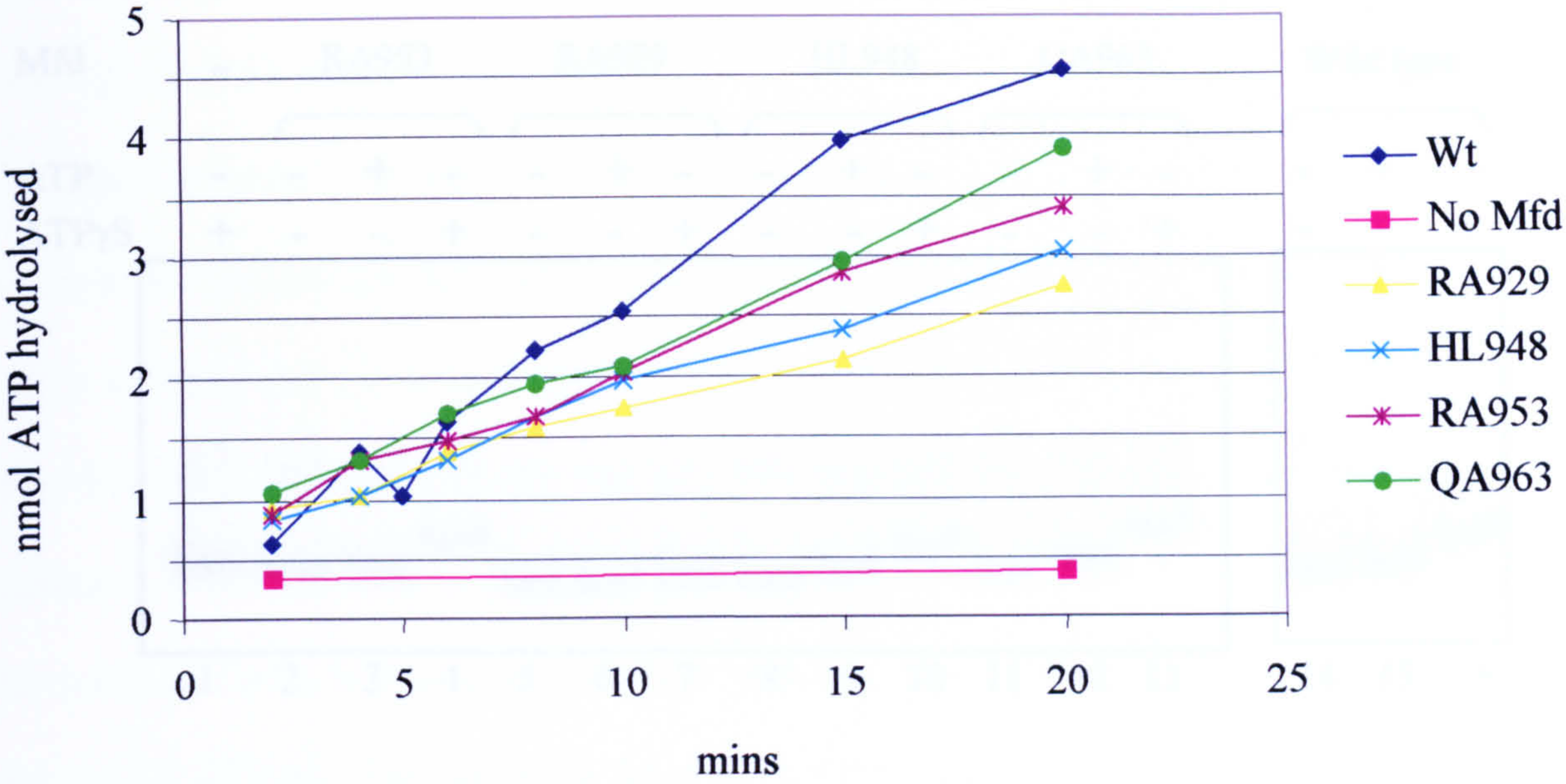
A k_{cat} value of 3 min^{-1} has been quoted for ATP hydrolysis by Mfd (Selby and Sancar, 1993a). To determine whether substitutions in the TRG motif that affected Mfd-mediated elongation complex displacement *in vivo* were defective in ATP binding or hydrolysis, ATPase assays were performed. k_{cat} values for the Mfd proteins containing substitutions in the TRG motif were determined in assays containing 1 μM Mfd protein (figure 4.7). The wild-type protein possessed a k_{cat}

value of 22 min^{-1}, which is higher than the previously quoted figure of 3 min^{-1} but is within experimental error at the low k_{cat} values involved. The four mutants all hydrolysed ATP with k_{cat} values slightly lower than the wild type but again within experimental error. Each of the substituted Mfd proteins was still able to bind and hydrolyse ATP suggesting that the fold of this region of the protein is not grossly affected by these substitutions. The failure of Mfd RA953 and Mfd QA963 to increase roadblock repression *in vivo* may not be due to an absence of ATP hydrolysis. However, it is possible that the ATPase activity of Mfd is stimulated *in vivo* and these mutants could be defective in stimulated ATP hydrolysis.

DNA binding activity of Mfd proteins containing TRG substitutions

Mfd-mediated elongation complex displacement requires 26 bp of accessible DNA upstream of the stalled RNA polymerase and therefore the defect in RNAP displacement *in vivo* by Mfd proteins containing substitutions in the TRG motif could be caused by a defect in DNA-binding activity. Mfd forms a stable complex with DNA in the presence of ATPγS but not in the presence of ATP. The DNA binding activity of each of the substituted Mfd proteins was tested under both conditions to determine whether it resembled that of the wild-type protein. The 250 bp *Eco*RI-*Bam*HI fragment from pSRc33 was end-labelled and was incubated with 100 nM Mfd proteins in the absence of nucleotide or in the presence of 2 mM ATP or 2 mM ATPγS. The complexes were analysed on a native polyacrylamide gel (figure 4.8).

In the absence of nucleotide, DNA that was incubated with wild-type Mfd and with each of the four mutant proteins ran with the same electrophoretic mobility as the DNA alone (lanes 1, 2, 5, 8, 11 and 14). The wild-type, RA929, HL948, RA953 and



Mfd protein	k_{cat}
Wild-type	22 +/- 3 min^{-1}
RA929	9.9 +/- 2.96 min^{-1}
HL948	12 +/- 0.25 min^{-1}
RA953	14 +/- 0.7 min^{-1}
QA963	15 +/- 3.39 min^{-1}

Figure 4.7. The ATPase activity of purified Mfd proteins containing single amino acid substitutions within the putative TRG motif of Mfd.

1 μM purified Mfd protein was incubated in a 10 μl reaction with 2mM ATP and 1 μCi [γ-^{32}P] ATP at 37°C. Unhydrolysed ATP was adsorbed using activated charcoal and Cerenkov counting was used to quantify hydrolysed ATP at different time points. The graph shown is the average of three experiments and k_{cat} values were calculated with standard deviation.

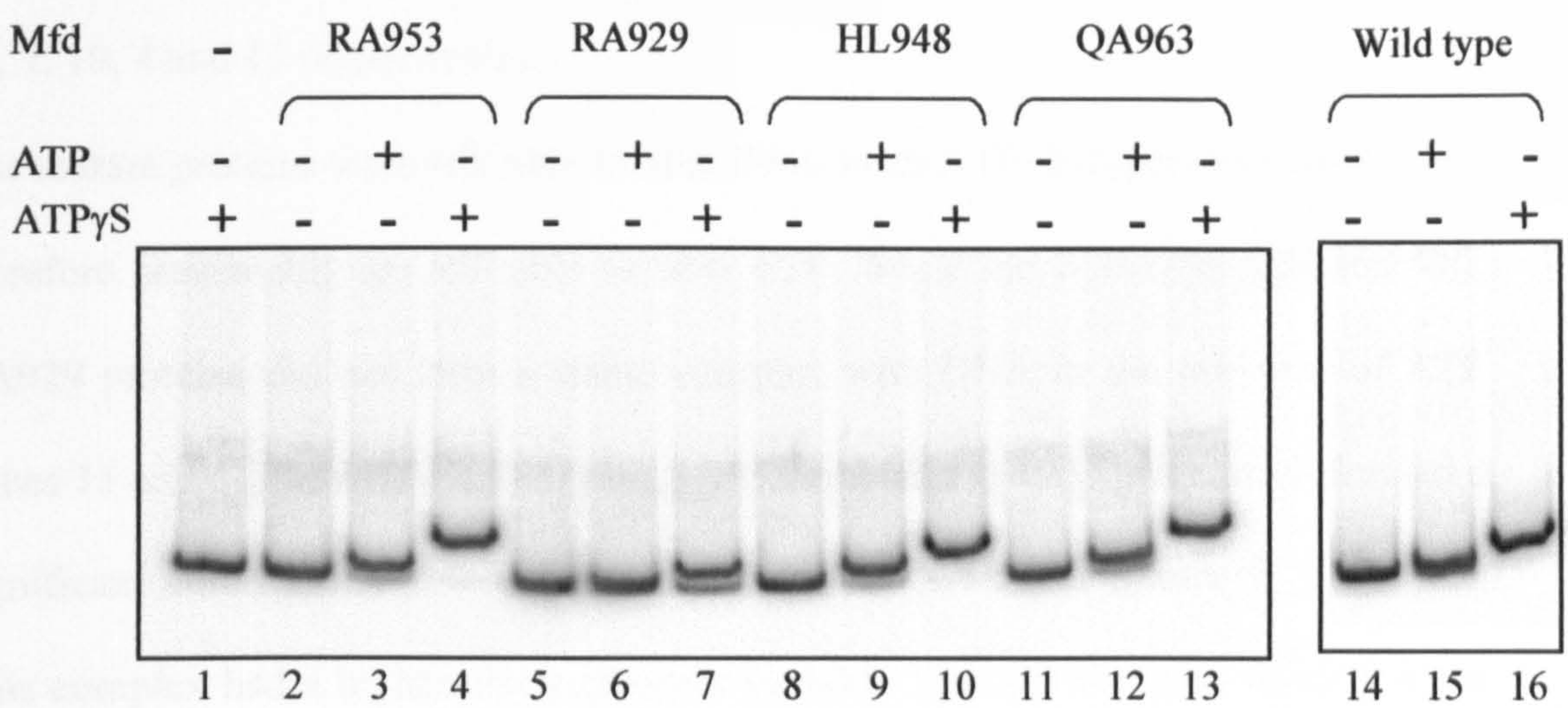


Figure 4.8. The effect of amino acid substitutions within the putative TRG motif of Mfd on DNA-binding *in vitro*.

Purified wild-type Mfd or Mfd carrying the indicated substitutions was incubated for 25 minutes at 37°C with an *Eco*RI-*Bam*HI end-labelled fragment from pSRc33. 2 mM ATP or ATPγS was included as indicated. The complexes were analysed using a 5% acrylamide/TAE, 8 mM magnesium acetate gel and a representative gel is shown.

QA963 Mfd proteins all formed a stable complex in the presence of ATPγS (lanes 16, 7, 10, 4 and 13 respectively).

The mutant proteins were still able to bind DNA in an ATPγS-dependent manner and therefore presumably are still able to bind ATP. While the wild-type Mfd and Mfd RA929 proteins did not form a stable complex with DNA in the presence of ATP (lanes 15 and 6), the Mfd HL948, Mfd RA953 and Mfd QA963 proteins all formed a significant amount of stable complex in the presence of ATP (lanes 9, 3 and 12). This complex had a higher electrophoretic mobility than the complex formed in the presence of ATPγS. The complex formed by Mfd HL948, Mfd RA593 and Mfd QA963 in the presence of ATP resembled the complex that Mfd RA929 formed in the presence of ATPγS.

The absence of a stable DNA-wild-type Mfd complex in the presence of ATP could be due to either ATP hydrolysis simply causing release of the DNA, or ATP hydrolysis powering translocation along and off the end of the DNA fragment. Therefore the presence of DNA-bound complex with Mfd HL948, Mfd RA953 and Mfd QA963 proteins in the presence of ATP may indicate a subtle defect in ATPase activity that was too small to be detected within the ATPase assay, a defect in coupling ATP hydrolysis to translocation along DNA or alternatively a defect in DNA release.

Bands with differing electrophoretic mobility were observed upon binding of wild-type Mfd and Mfd RA929 to a 250 bp DNA fragment in the presence of ATPγS. In order to further investigate the nature of these different complexes , which possessed different electrophoretic mobilities further, titrations of wild-type Mfd and Mfd proteins containing substitutions in the TRG motif were performed on the DNA fragment used above.

Purified wild-type Mfd, MBP-Mfd and Mfd proteins containing substitutions in the TRG motif were titrated into a reaction containing the 250 bp *Eco*RI-*Bam*HI fragment of pSRc33 and ATPγS. The resulting complexes were analysed by EMSA (figure 4.9). At 50 nM each of the proteins caused some retardation of the DNA fragment (compare lane 1 to lanes 2, 6, 10, 14 and 18). Whilst the MBP-Mfd, wild-type Mfd, Mfd RA953 and Mfd HL948 proteins bound all of the DNA at 50 nM protein, at this concentration the Mfd RA929 protein did not retard all of the fragment. This may suggest that Mfd RA929 possesses a lower affinity for DNA. Addition of 200 nM Mfd protein (lanes 3, 7, 11, 15 and 19) resulted in the formation of complexes with lower electrophoretic mobility than the complexes that were formed when 50 nM protein was added. At 350 nM and 500 nM the Mfd HL948 and Mfd RA953 proteins formed complexes with a much lower electrophoretic mobility than the complex formed at equivalent concentrations of wild-type Mfd protein. The dramatically retarded species observed at high concentrations of Mfd HL948 and Mfd RA953 may indicate formation of higher order structures of Mfd bound to DNA, and may be due to an increased binding affinity of these proteins for DNA.

The change in electrophoretic mobility of the wild-type Mfd-DNA complex as the Mfd protein was titrated into the reaction may be the result of multiple molecules of Mfd binding to one DNA fragment. This may indicate that the complex formed by 100 nM Mfd RA929 in the presence of ATPγS contains fewer Mfd molecules bound to the DNA fragment than the complex formed by the wild-type protein under the same conditions, again suggesting that the purified Mfd RA929 possesses lower DNA binding affinity. Binding of the 130 kDa Mfd protein to a 250 bp double-stranded DNA fragment produced a relatively small change in the electrophoretic

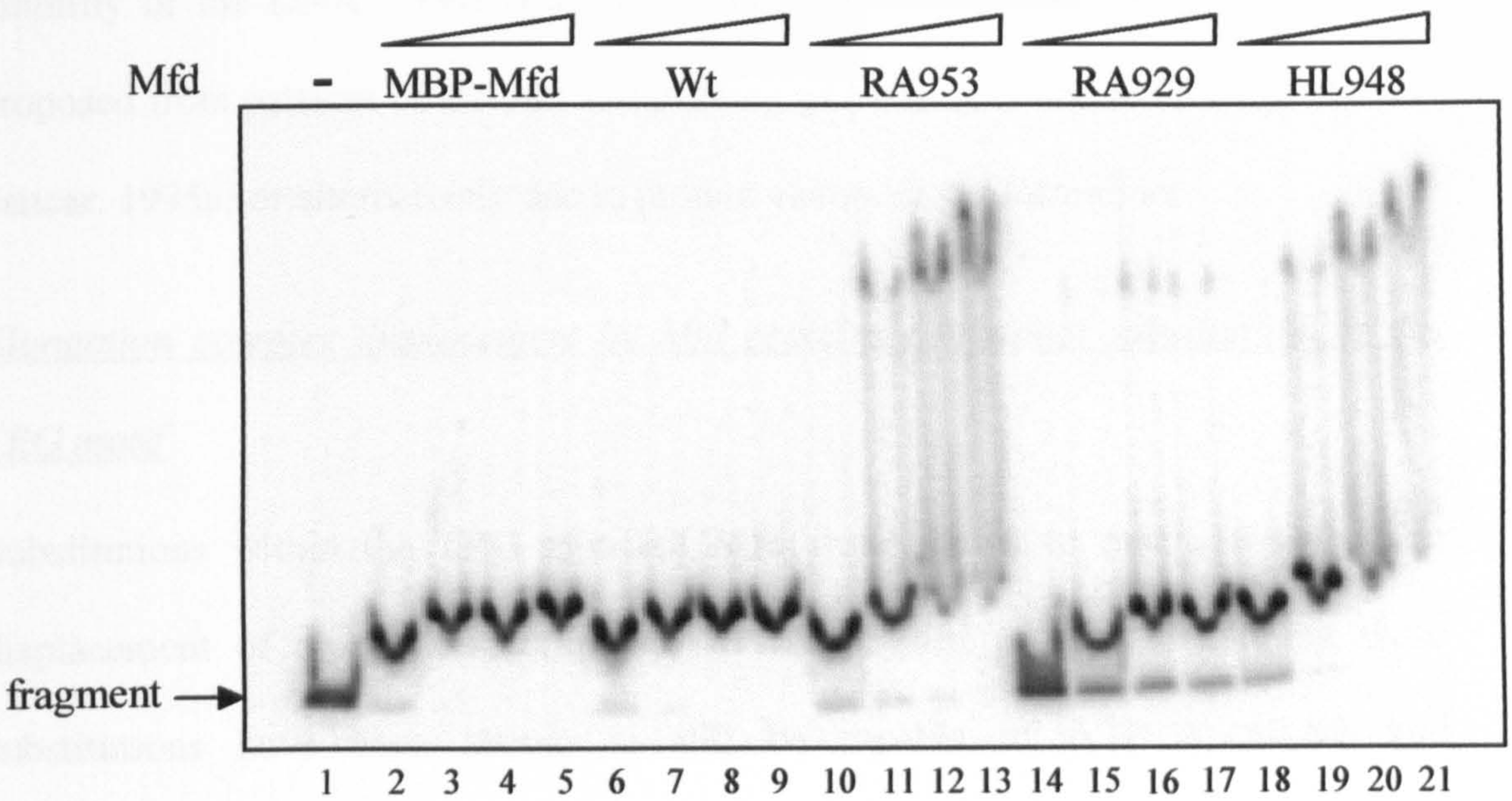


Figure 4.9. Titration of Mfd proteins into DNA-binding assays in the presence of ATPγS.

0.4 nM 250 bp *Eco*RI-*Bam*HI of pSRc33 was incubated with 50 nM (lanes 2, 6, 10, 14 and 18), 200 nM (lanes 3, 7, 11, 15 and 19), 350 nM (lanes 4, 8, 12, 16 and 20) and 500 nM (lanes 5, 9, 13, 17 and 21) purified Mfd protein in the presence of 2 mM ATPγS. The complexes were analysed using a 5% acrylamide/TAE, 8 mM magnesium acetate gel.

mobility of the DNA. This could be a consequence of DNA wrapping that was proposed from patterns of DNaseI footprinting of Mfd_{trunc} bound to DNA (Selby and Sancar, 1995a) or alternatively due to protein charge or conformation.

Elongation complex displacement by Mfd proteins containing substitutions in the TRG motif

Substitutions within the TRG motif of Mfd were shown to result in defects in displacement of elongation complexes *in vivo*. Mfd proteins containing these substitutions have been shown to still be capable of ATP hydrolysis and ATPγS-dependent DNA binding *in vitro* and therefore it is possible that these proteins are specifically defective in elongation complex displacement or coupling of ATP hydrolysis to displacement. Mfd proteins containing substitutions in the putative TRG motif were tested for the ability dissociate elongation complexes stalled by nucleotide starvation *in vitro*.

Elongation complex displacement was assayed by EMSA. The *Rsa*I-*Sma*I fragment from pAR1707 contains the T7 A1 promoter and transcription complexes initiating from this promoter are stalled at +20 in the absence of UTP when incubated with a transcription reaction mixture containing GTP, ATP, CTP and ApU as an initiating dinucleotide (figure 4.10 A). This fragment was end-labelled with ^{32}P to permit analysis of unbound DNA, initiation complexes and elongation complexes (figure 4.10 B). Initiation complexes were distinguished from stalled elongation complexes based on their differing electrophoretic mobilities (compare lanes 2 and 4). Labelled fragment at 0.4 nM was incubated with 5 nM RNAP to form initiation complexes and unstable complexes were removed by the addition of heparin.

A

B

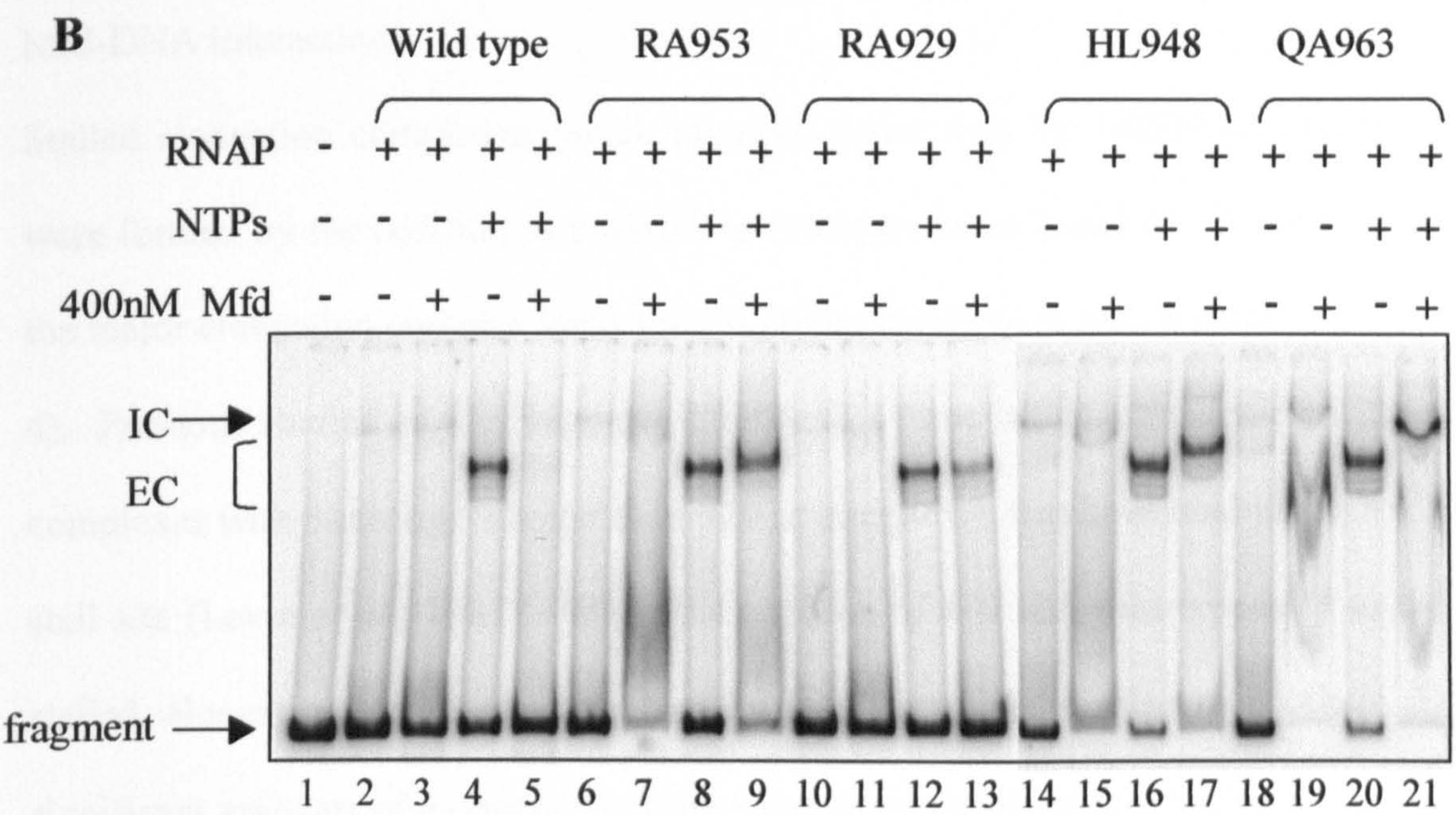


Figure 4.10. The effect of amino acid substitutions within the putative TRG motif of Mfd on Mfd-mediated *in vitro* displacement of RNAP: analysis using labelled template.

(A) The sequence of the transcript produced from the *Rsa*I-*Sma*I fragment of pAR1707. (B) Initiation complexes were formed by incubation of a ^{32}P end-labelled *Rsa*I-*Sma*I fragment from pAR1707 with RNAP holoenzyme. Elongation complexes were formed at +20 by the addition of ApU, ATP, CTP and GTP (NTPs). Purified wild-type Mfd protein or purified protein containing the indicated amino acid substitution was added where indicated. Complexes were analysed using a 4.5% acrylamide EMSA gel and the bands corresponding to initiation complexes (IC), elongation complexes (EC) and the unbound fragment are indicated. A representative gel is shown.

The initiation complexes were not very stable under the electrophoresis conditions that were employed (lanes 2, 6, 10, 14 and 18). Incubation with 400 nM of each of the Mfd proteins did not result in dissociation of the initiation complex (lanes 3, 7, 11, 15 and 19), although the Mfd HL948, Mfd RA953 and Mfd QA963 proteins bound non-specifically to the DNA to cause shifting and smearing of the free fragment (lanes 15, 7 and 19). The observed shift of the DNA may be solely due to Mfd-DNA interactions.

Stalled elongation complexes, which migrated faster than the initiation complexes were formed by the addition of nucleotides (compare lanes 2 and 4). In addition to the major elongation complex band several fainter bands were also present (e.g. lane 4). Previous studies on this fragment indicate that these bands represent elongation complexes with transcripts shorter than +20 or complexes that have read-through the stall site (Levin *et al.*, 1987). While the addition of 400 nM wild-type Mfd to the stalled elongation complexes resulted in their complete displacement (lane 5) significant amounts of elongation complex remained following incubation with each of the mutant proteins (lanes 9, 13, 17 and 21).

Therefore RA929, HL948, RA953 and QA963 substitutions impair the ability of Mfd to displace elongation complexes. However, it is possible that the substituted proteins caused dissociation of the transcript from stalled elongation complexes but failed to remove the RNA polymerase protein from the DNA. In order to determine whether the complexes that remain following incubation with the mutant Mfd proteins are functional elongation complexes, containing nascent RNA, the elongation complexes were also analysed by incorporation of label into the RNA.

Once again the *Rsa*I-*Sma*I fragment from pAR1707 was used as a template and stalled elongation complexes were formed by incubation of the fragment with

holoenzyme followed by the addition of heparin, ApU, ATP, GTP and CTP along with α^{32}P-labelled CTP. The stalled complexes were then incubated with 400 nM Mfd and were analysed by EMSA (figure 4.11). Addition of wild-type Mfd caused the displacement of the elongation complex (compare lanes 1 and 2).

Some elongation complex displacement occurred following incubation with Mfd RA929 (lane 4), but only a fraction of that observed with the wild-type protein. Mfd RA953 and Mfd QA963 'supershifted' the complex i.e. decreased the electrophoretic mobility of the stalled elongation complex (compare lanes 2, 3 and 6) and indeed 'supershifting' was observed for QA963 in figure 4.10 (lane 21). Therefore RA953 and QA963 have lost the ability to displace elongation complexes but are still able to bind to RNAP or DNA. A 'supershift' is also observed on addition of HL948 (figure 4.10 lane 17 and figure 4.11 lane 5).

The EMSAs described above analysed elongation complex displacement at a single time point, 15 minutes after addition of Mfd to stalled elongation complexes and showed that the substituted proteins were defective in RNAP dissociation. The substituted Mfd proteins may either be completely unable to displace stalled elongation complexes or may possess a defect that dramatically reduces the rate of RNAP dissociation. In order to distinguish between these possibilities and to determine the rate of displacement by each of the mutants, time courses following elongation complex displacement by each of the purified Mfd proteins were performed by Dr. A. Smith.

Stalled elongation complexes on labelled DNA were formed as above. Mfd protein was added at 250 nM and samples were taken and run on EMSA gels at different time points (figure 4.12). In the absence of Mfd the elongation complex alone was stable throughout the time course of the experiment. More than 90% of elongation

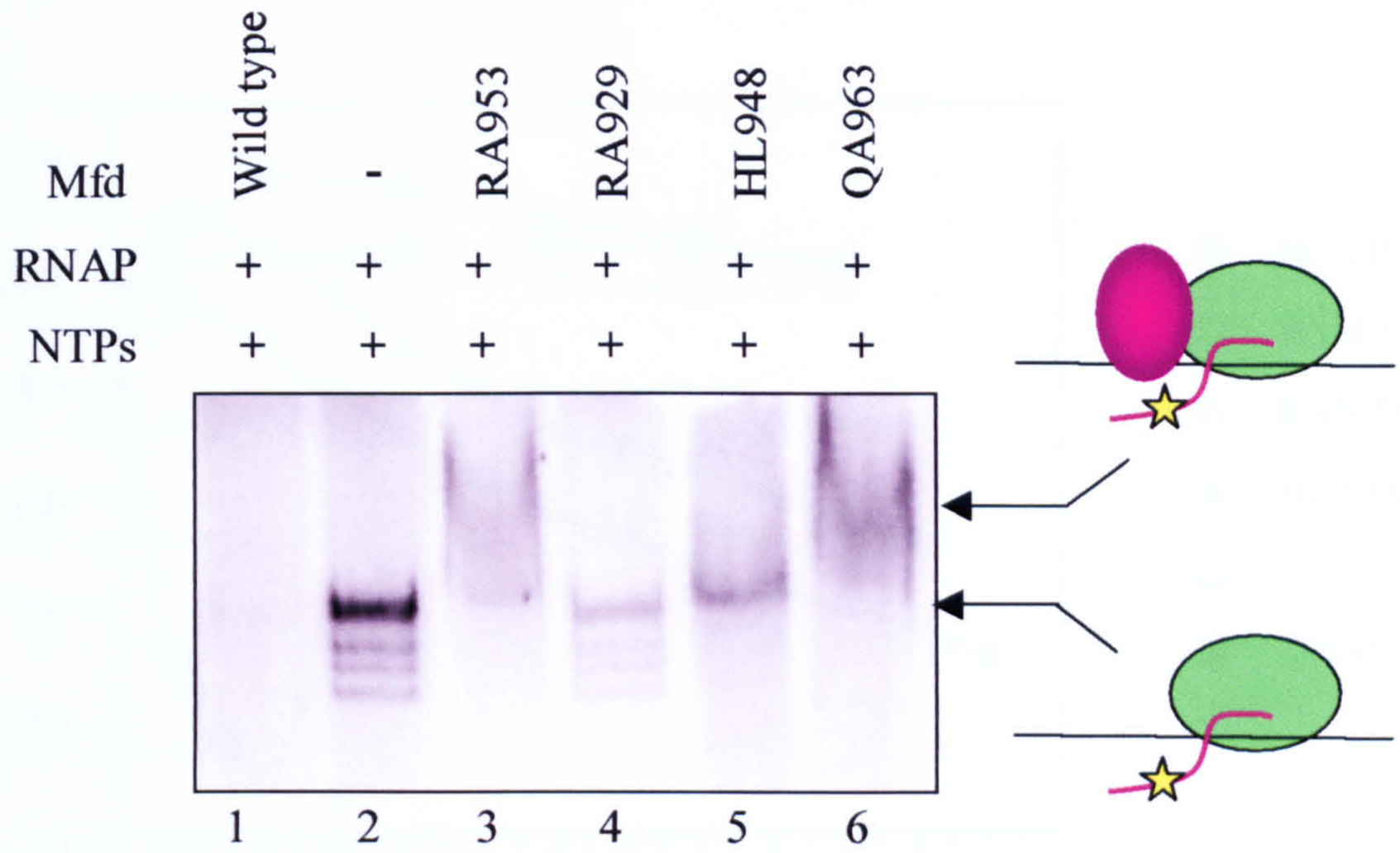


Figure 4.11. The effect of amino acid substitutions within the putative TRG motif of Mfd on Mfd-mediated *in vitro* displacement of RNAP: analysis using labelled nascent transcript.

Initiation complexes were formed by incubation of an *Rsa*I-*Sma*I fragment from pAR1707 with RNAP holoenzyme. Elongation complexes in which the nascent transcript was labelled by incorporation of [α-^{32}P] CTP were formed by the addition of ApU, ATP, CTP and GTP (NTPs). Purified wild-type Mfd protein or purified protein containing the indicated amino acid substitution was added where indicated. Complexes were analysed using a 4.5% acrylamide EMSA gel and a representative gel is shown.

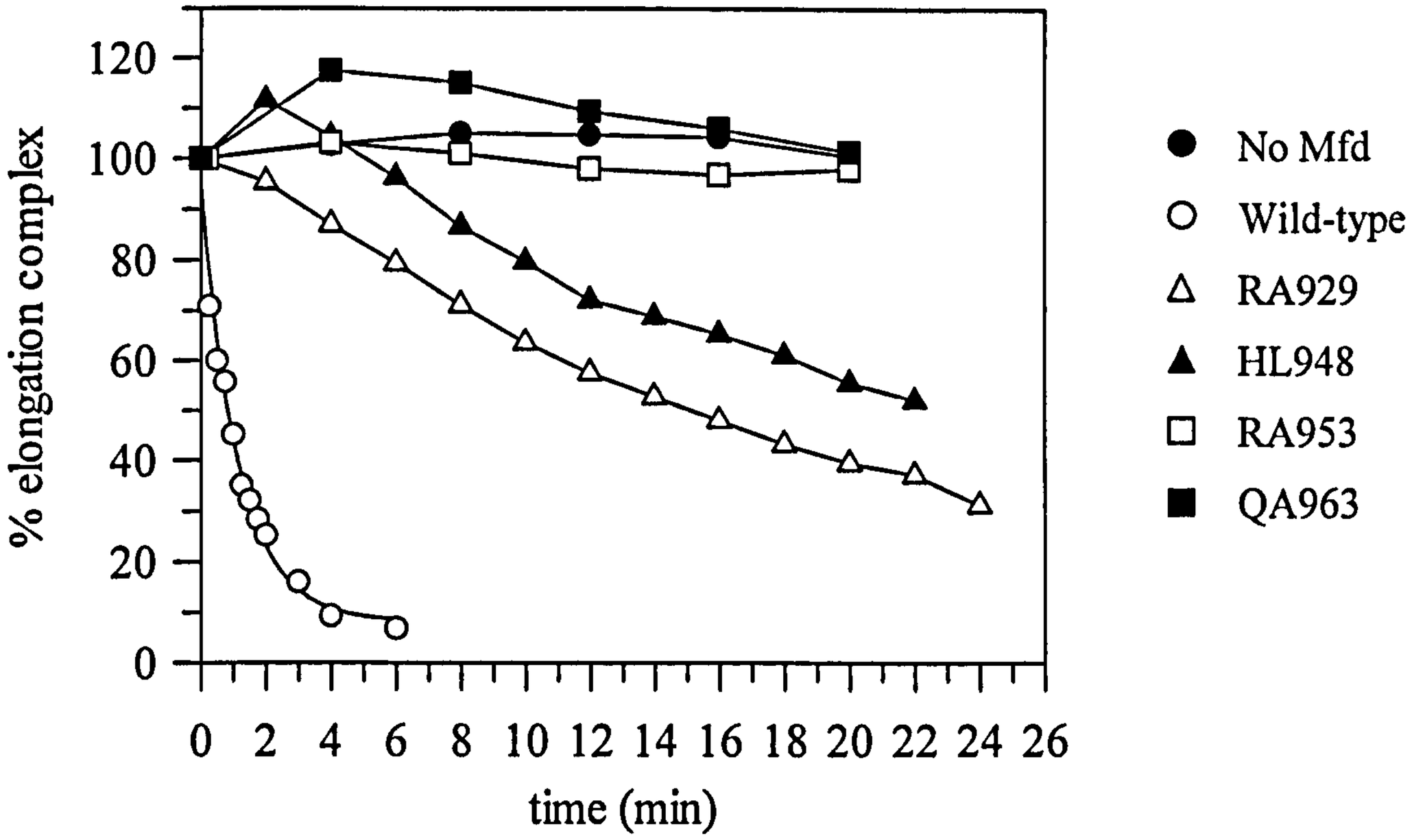


Figure 4.12. Time course of elongation complex displacement by Mfd proteins.

Elongation complexes stalled by nucleotide starvation on the *Rsa*I-*Sma*I fragment of pAR1707 were formed. Their displacement at 37°C over time in the absence of Mfd (filled circles) and in the presence of purified wild-type Mfd protein (open circles) and proteins containing RA929 (open triangles), HL948 (filled triangles), RA953 (open squares) and QA963 (filled squares) substitutions was analysed by EMSA. Values are the average of three experiments and are expressed as a percentage of the elongation complex present at time 0.

complexes were displaced following a 5 minute incubation with wild-type Mfd. In contrast Mfd RA953 and Mfd QA963 proteins did not result in a decrease in the amount of elongation complex, even after 20 minutes. Mfd RA929 and Mfd HL948 proteins were still able to displace elongation complexes but were considerably slower the wild-type Mfd protein.

On the basis of their *in vitro* properties the substitutions can be split into two classes. Mfd proteins containing RA953 and QA963 substitutions are unable to increase the efficiency of roadblock repression *in vivo* and *in vitro* and are unable to displace elongation complexes, despite still binding to RNAP. Mfd RA929 and Mfd HL948 proteins are able to complement an *mfd*⁻ phenotype *in vivo* in liquid culture and are still able to displace elongation complexes *in vitro* but much more slowly than the wild-type protein. Therefore R953 and Q963 are essential residues for elongation complex displacement and we can conclude that while R929 and H948 contribute to RNAP displacement, they are not essential.

DISCUSSION

Four single amino acid substitutions, QL898, HL948, GD956 and GV960 were isolated from an *in vivo* screen to identify *mfd* mutants that fail to increase roadblock repression efficiency. Q898 is a conserved residue within helicase motif VI, while H948, G956 and G960 are located in a region downstream of the helicase motifs, which shows a high degree of sequence similarity to the TRG motif of RecG. In addition the highly conserved R929, R953 and Q963 residues, which had been shown to be essential for DNA unwinding in RecG, were substituted with alanine by site-directed mutagenesis.

Plasmids encoding these seven substitutions were tested for their ability to increase roadblock repression *in vivo*. On solid media each of the substitutions conferred chloramphenicol-resistance to *mfd*$^-$ cells containing the roadblock reporter construct pRCB-CAT, indicating that the mutants had lost the ability to enhance repression of genes downstream of a roadblock. The simplest explanation of these data is that the Mfd proteins are defective in displacement of the transcription complex that is stalled at the roadblock. In roadblock repression assays performed in liquid culture cells containing Mfd QL898, Mfd RA953, Mfd GD956, Mfd GV960 and Mfd QA963 proteins had a level of CAT activity equivalent to cells lacking the Mfd protein. However, Mfd RA929 and Mfd HL948 were still able to increase the efficiency of roadblock repression in liquid culture. This discrepancy between the effects seen on solid media and in liquid culture may be due to the difference in physiology between cells within a colony on an agar plate and in liquid media, such as a difference in pH, anaerobic vs. aerobic conditions or cell density. The subsequent discovery that purified Mfd RA929 and Mfd HL948 proteins are

defective in elongation complex displacement *in vitro* indicates that the phenotype on chloramphenicol agar plates may be the more sensitive assay of Mfd function.

Mfd proteins encoding RA929, HL948, RA953 and QA963 substitutions were purified and their *in vitro* properties were tested. All four of these proteins were still able to bind and hydrolyse ATP. The wild-type Mfd protein binds dsDNA when bound to ATP but upon ATP hydrolysis Mfd dissociates from the DNA. The loss of Mfd from DNA could either occur simply by release upon ATP hydrolysis or could be due translocation along and off the end of the linear DNA. It may be interesting to see whether in the presence of ATP, Mfd is released from minicircle DNA or from linear DNA with a blockage to translocation. In addition one could examine the effect that analogues that mimic different stages of ATP hydrolysis (e.g. ADP, AMPPNP, $ADPBeF_3$ and $ADPAlF_4^-$) have on DNA binding by Mfd since movement through different stages of the ATP hydrolysis cycle is associated with protein conformational changes.

Titration of wild-type Mfd protein into a DNA binding assay containing a fragment 250 bp long resulted in the formation of complexes with greater electrophoretic mobility at low Mfd concentrations than at higher Mfd concentrations. The differing migration of these complexes could be due to the binding of different numbers of Mfd molecules to the DNA fragment or may indicate a difference in the conformations of the complexes, such as in the DNA bend angle since bending of DNA has been associated with increased retardation in EMSAs (Wu and Crothers, 1984).

All four of the purified Mfd proteins containing substitutions within the TRG motif were still able to bind to DNA. Mfd HL948, Mfd RA953 and Mfd QA963 proteins formed a complex with DNA in the presence of ATP with a higher electrophoretic

mobility than that of complexes formed in the presence of ATPγS. This may indicate that either the proteins have a subtle defect in ATP hydrolysis or that they possess a defect in coupling ATP hydrolysis to DNA release or translocation. At high concentrations of Mfd RA953 and Mfd HL948 some Mfd-DNA complexes with greatly reduced mobility were observed. These higher order complexes may indicate an increased affinity of these two substituted proteins for DNA or may be a result of protein-protein contacts.

It is interesting to note that the complex Mfd RA929 formed in the presence of ATPγS resembled that formed by the other mutants in the presence of ATP. Therefore substitution of R929 may prevent the Mfd protein adopting the same conformation upon ATP binding that the other mutants do. Substitution of R929 may place the DNA binding region of Mfd in a conformation that resembles the conformation adopted by the other mutants when they begin ATP hydrolysis. Alternatively, the fact that complexes formed by 100 nM Mfd RA929 in the presence of ATPγS migrated to the same position as complexes formed by 50 nM wild-type Mfd protein in the presence of ATPγS, and that a higher concentration of Mfd RA929 was required to fully retard the fragment, may indicate that the purified Mfd RA929 possesses a slightly reduced affinity for DNA.

Mfd RA929, Mfd HL948, Mfd RA953 and Mfd QA963 proteins are all defective in elongation complex displacement *in vitro*. These substitutions fall into two groups; the first containing RA953 and QA963 in which elongation complex dissociation has been abolished and the second containing RA929 and HL948, which are still able to displace elongation complexes but at a much reduced rate. These two groups are the same as those established in liquid culture roadblock repression assays where Mfd RA929 and Mfd HL948 encoding plasmids were able to increase the efficiency of

roadblock repression and Mfd RA953 and Mfd QA963 behaved as if there was no Mfd present. Residues R953 and Q963 therefore are essential for elongation complex displacement while R929 and H948 may contribute, but are not absolutely required. The strength of the effect of substitution of each of the two arginine residues of the helical hairpin of the TRG motif echoes the hierarchy of the effect of the equivalent substitutions in RecG. R609 in RecG is equivalent to R929 in Mfd and substitution of this arginine had more modest effects both *in vivo* and on *in vitro* DNA unwinding activity than substitution of R630 (equivalent to R953 in Mfd).

Substitutions within the putative TRG motif of Mfd have been found to affect the ability of Mfd to displace stalled elongation complexes both *in vivo* and *in vitro*, demonstrating that a functional TRG motif is required for Mfd function. A model of the helicase domains and TRG motif of Mfd was constructed based on homology modelling to RecG (figure 4.13 (A)). The residues substituted by site-directed mutagenesis and those isolated from random mutagenesis screening are shown in figure 4.13 (B).

$MfdTrunc_{1\text{-}939}$, which is defective in elongation complex displacement *in vitro*, was also defective in RNAP dissociation *in vivo*. In contrast, $MfdTrunc_{1\text{-}997}$ retained much of the ability of wild-type Mfd to increase roadblock repression *in vivo* and recent preliminary experiments in our lab. have indicated that it is also able to displace elongation complexes stalled by nucleotide starvation *in vitro*. One conclusion that can be drawn is that the region between amino acid 939 and amino acid 997, which includes the TRG motif, is required for Mfd- mediated displacement of stalled elongation complexes *in vivo* and the C-terminal 151 amino acids of Mfd are not essential for this function. The structural model explains the properties of the

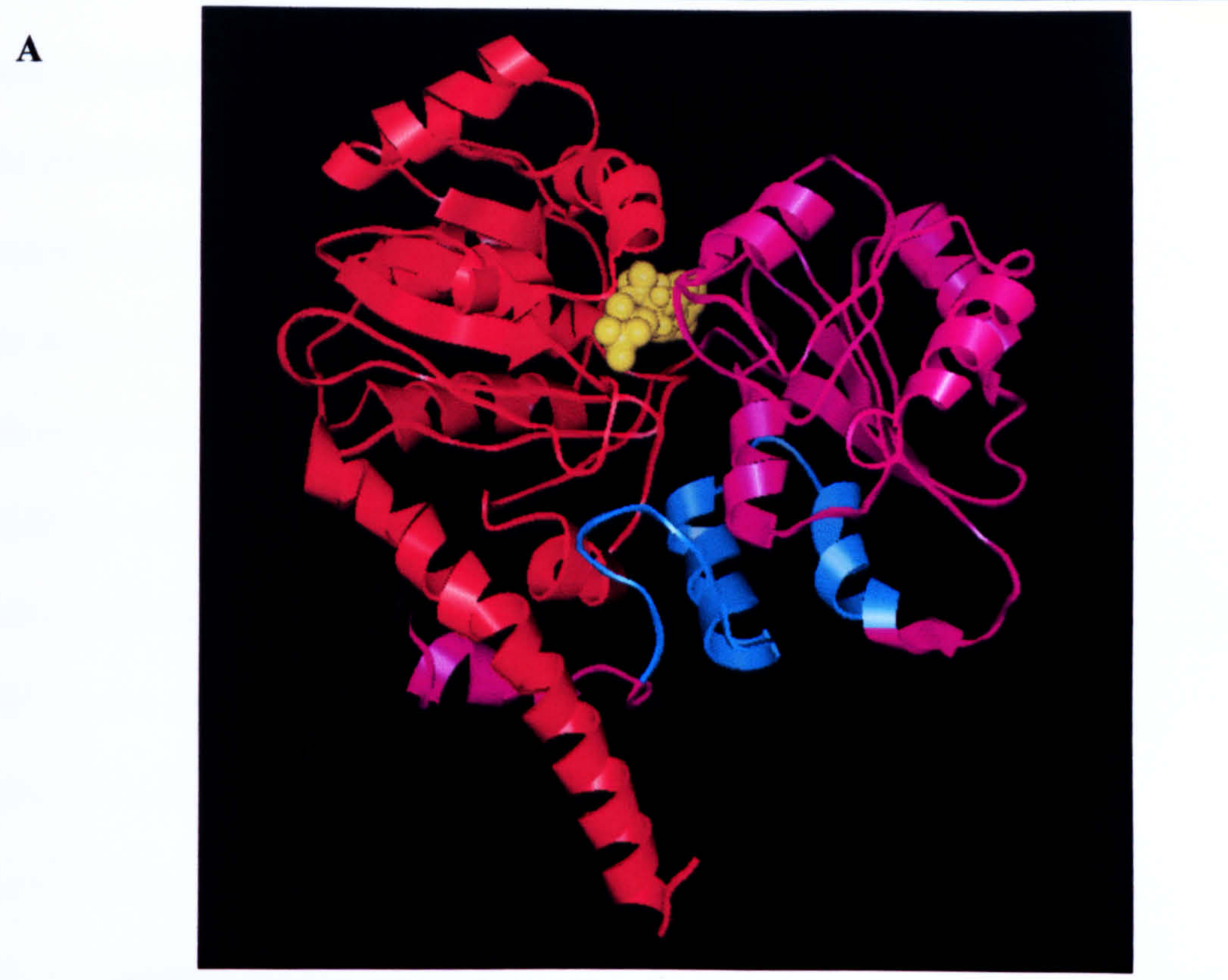


Figure 4.13. Homology model of the helicase domains of Mfd.

(A) Residues 541-983 of *E. coli* Mfd were aligned with and modelled onto residues 309-755 of *T. maritima* RecG using ClustalW and InsightII 97. The two helicase domains are shown in red and magenta, ADP in yellow and the TRG motif in cyan. (B) A close-up of motif VI and the TRG motif of Mfd showing the position of residues altered by random (yellow) and site-directed mutagenesis (green).

$MfdTrunc_{1-939}$ and $MfdTrunc_{1-997}$ proteins. Whilst $MfdTrunc_{1-997}$, which retained the ability to increase roadblock repression contains an intact TRG motif, $MfdTrunc_{1-939}$ is truncated within the loop between the two helices of the TRG hairpin and therefore lacks a functional TRG motif.

A model was proposed for the coupling of ATP hydrolysis in RecG to translocation along dsDNA via the TRG motif (Mahdi *et al.*, 2003). The model involves transmission of conformational changes from the site of ATP hydrolysis, through motif VI to the TRG motif. This model for coupling ATP hydrolysis to a conformational change in the TRG motif is plausible for Mfd. Q898 is located within motif VI in an α-helix, which runs parallel to the helices of the TRG hairpin. Motif VI has been implicated in coupling ATP hydrolysis to helicase activity (Hall and Matson, 1999). A conserved arginine within motif VI forms a salt bridge with the γ phosphate of ATP and within superfamily 2 helicases the glutamine equivalent to Q898 interacts with the histidine of motif II. In DEAD-box helicases a histidine located in this position in motif VI interacts with an aspartate of motif II. Substitution of the equivalent of Q898 in CSB resulted in a protein that was impaired in both ATP hydrolysis and ATP binding (Christiansen *et al.*, 2003) and in NS3 was shown to abolish ATPase and helicase activity completely (Tai *et al.*, 2001). Therefore ATP hydrolysis may lead to the movement of motif VI and therefore the movement of the second helicase domain in relation to the first helicase domain. Substitution of Q898 may abolish the coupling of ATP hydrolysis to motif VI and therefore prevent coupling to the TRG motif.

A conserved glutamate located at the N-terminus of the helix that includes motif VI may stabilise the two arginine residues within the helical hairpin via a hydrogen bond network. When this residue was substituted in RecG the protein was defective

in DNA unwinding *in vitro* and the substitution shows reduced complementation in *recG*⁻ cells when combined with R609 substitutions. Therefore ATP hydrolysis may result in movement of motif VI, which may disrupt the hydrogen bond network between D889 and the TRG motif. This may cause a change in the relative orientation of the two helices of the hairpin that could be transmitted to the remainder of the motif. Substitution of either R929 or R953 would therefore disrupt the hydrogen bond network and affect conformational changes occurring in this region of the protein following ATP hydrolysis. The effect of substitution of H948 may either be direct since it removes an ionisable side chain from the immediate vicinity of the R929 and R953 or indirect, since it neighbours a conserved aspartate residue that is within hydrogen-bonding distance of these arginine residues. The importance of having a seven amino acid loop between the helices of the TRG helical hairpin in Mfd compared to only four residues in RecG is unclear but it acts as a signature for the TRG motifs of the two proteins.

The TRG motif of RecG has been implicated in translocation along double-stranded DNA. In order for a monomeric protein to translocate along DNA it must possess two DNA contact sites. The two sites act in conjunction, alternating binding and release of DNA in an inchworm mechanism. Alternatively one of the sites may be replaced by a contact with another DNA bound protein as occurs for the chromatin remodelling protein, ISW (Fitzgerald *et al.*, 2004). Since the path of the DNA across the helicase domains and TRG motif of RecG is unknown the manner in which a change in conformation of the TRG motif causes translocation remains speculative. It was proposed that the function of the loop that follows the helical hairpin in RecG is as a DNA-binding arm and that the conserved glutamine within this loop may act as a DNA binding residue. The loop was postulated to either act as a DNA lever that

moves following ATP binding and hydrolysis or alternatively as a ratchet to catch the DNA following translocation by a different mechanism. Substitution of Q963 within this loop in Mfd results in a protein that is unable to displace elongation complexes but that is still able to bind DNA. This does not necessarily mean that this residue does not contact DNA but only that it is not essential for the non-specific binding of Mfd to dsDNA. It may be that this glutamine residue does form a contact with DNA that is only required during translocation along DNA and that a second DNA binding site is present in the Mfd protein.

However in order to act as a DNA lever one would expect this loop region to be rigid. Whilst in RecG the loop contains a conserved proline residue that would provide some rigidity the situation is different in Mfd. In Mfd sequences, not only does a small aliphatic residue replace this proline but across Mfd and RecG sequences the loop also contains four highly conserved glycine residues, providing conformational flexibility. Substitution of two of these glycine residues (G956 and G960) resulted in Mfd proteins that were unable to increase the efficiency of roadblock repression *in vivo*. These substitutions would be expected to decrease the flexibility of the loop and imply that in fact flexibility of this loop is required for Mfd function rather than it acting as a stiff DNA-binding arm. Additionally, in *E. coli* Mfd the loop contains three negatively charged glutamate residues, an unlikely sequence for a region proposed to contact the negatively charged phosphates of the DNA backbone. It is clear however that the loop of the TRG motif is essential for elongation complex displacement by Mfd.

Given that the TRG motif of RecG has been shown to be required for branch migration of Holliday junctions and that Mfd acts on DNA-RNAP complexes, it seems likely that this motif is required for coupling ATP binding and hydrolysis to

translocation along dsDNA in both proteins by a common mechanism. RecG and Mfd appear to share a common 'motor' domain (although as yet there is no experimental evidence for the translocation of Mfd along naked DNA) and utilise additional domains to confer their specific function. In Mfd these include an RNAP interaction domain while in RecG it is a domain to separate the strands of a DNA junction. Other superfamily 2 helicases such as chromatin remodelling enzymes and type I restriction enzymes also translocate along dsDNA (Firman and Szczelkun, 2000; Whitehouse *et al.*, 2003) and it may be that despite no obvious sequence similarity to the TRG motif they possess a region that is structurally similar.

CHAPTER 5

THE DNA-BINDING AND NUCLEOTIDE-HYDROLYSIS PROPERTIES OF Mfd

DNA-binding properties of the Mfd protein

The DNA binding activity of Mfd was localised to amino acids 571-939 (Selby and Sancar, 1995a), corresponding to the helicase domains, but the path of the DNA across these domains is unknown. The helicase domains of Mfd show a high degree of similarity to those in RecG, however, the double stranded arm of the synthetic junction present in the crystal structure of RecG was not of sufficient length to determine the route of the DNA in this complex (Singleton *et al.*, 2001).

There is a requirement for DNA upstream of a stalled elongation complex for RNAP displacement by Mfd and there is a requirement for 90 bp of downstream DNA for transcription-coupled repair. Therefore, a greater understanding of the DNA-binding properties of Mfd may provide clues as to the mechanism of elongation complex displacement and other steps of transcription coupled-repair. The binding of Mfd to DNA of different lengths and the footprinting pattern of Mfd was examined *in vitro*.

Length dependence of DNA binding by Mfd

Mfd has a higher affinity for substrates that contain a region of approximately 90 bp of double-stranded DNA than for single-stranded DNA or RNA (Selby and Sancar, 1995a). Previous double-stranded DNA-binding assays using the *E. coli* Mfd protein have been carried out using fragments containing a minimum of approximately 90 bp of double-stranded DNA and yet only 26 bp of double-stranded DNA is required upstream of a stalled elongation complex for its displacement by Mfd. One would expect that the extent of any wrapping of DNA around the Mfd protein would be less for a short section of DNA and as such the binding affinity may be reduced in comparison to longer DNA. Therefore it is of interest in terms of the mechanism of

RNAP displacement to determine whether the DNA binding properties of Mfd are the same for short and long fragments of DNA.

In order to compare the affinity of Mfd for short and long DNA fragments, both a 32 bp double-stranded oligomer and a 250 bp fragment were incubated with wild-type Mfd protein in the presence of ATPγS and the resulting complexes were analysed by EMSA (figure 5.1). The 250 bp fragment was completely bound, forming a complex with decreased electrophoretic mobility, upon the addition of 50 nM Mfd (figure 5.1 A lanes 1 and 2). In contrast the addition of 50 nM Mfd to the 32-mer resulted in the formation of only a very modest amount of retarded complex (figure 5.1 B lanes 2 and 3). Indeed, even in the presence of 500 nM Mfd some free 32-mer remained (figure 5.1 B lane 7). Approximately 50% of the 32 bp DNA was bound in the presence of 200 nM protein (figure 5.1 B lane 5). The binding of Mfd to the 32-mer was dependent on the presence of ATPγS, as observed previously for the 250 bp fragment (figure 5.1 B compare lanes 1 and 7).

Binding of Mfd to the 250 bp fragment in an ATPγS-dependent manner was stronger than to the 32 bp oligomer of dsDNA. This could be a consequence of wrapping of the longer DNA fragment around the protein. It is interesting to note that the change in electrophoretic mobility upon Mfd binding is still relatively modest even on a 32-mer. Since there is a limit to the extent of wrapping that can occur for a 32 bp oligomer, the modest retardation of the DNA by Mfd may be due to the charge on the Mfd protein. Overall the Mfd protein contains 162 negatively charged amino acids (aspartate and glutamate) and 136 positively charged amino acids (lysine and arginine) resulting in a net charge of –26.

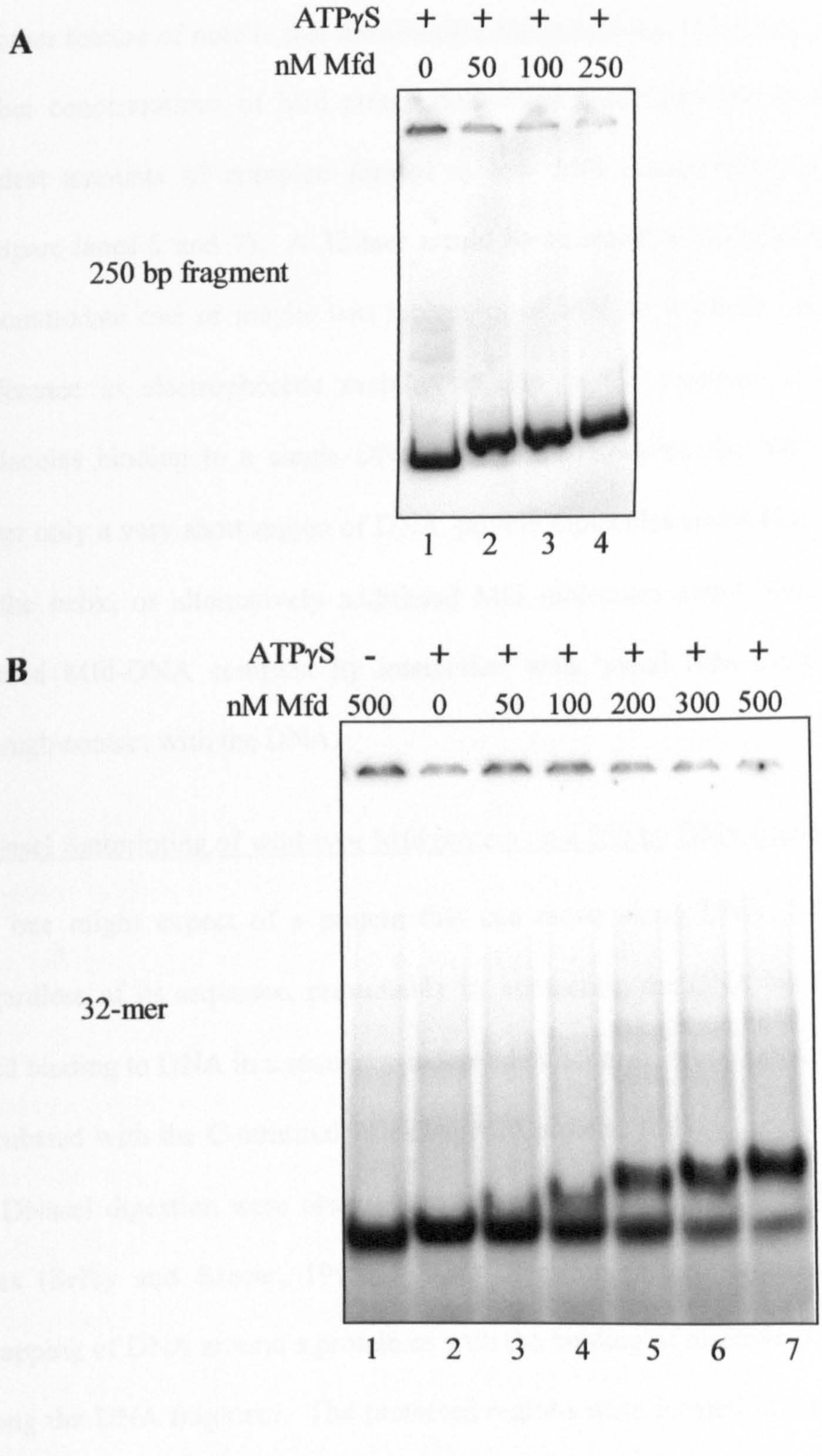


Figure 5.1. Binding of Mfd to 250 bp and 32 bp double-stranded DNA.

Wild-type Mfd protein was incubated at the concentration indicated with 2 mM ATPγS and (A) 0.4 nM *Eco*RI-*Bam*HI fragment from pSRc33 or (B) 0.4 nM 32 bp oligomer formed by annealing the oligonucleotides MFD012 and MFD013. Complexes were analysed by EMSA using an 8 mM magnesium acetate, 5% acrylamide/TAE gel and representative gels are shown.

Another feature of note is that the complex formed on the 32-mer in the presence of higher concentrations of Mfd protein had lower electrophoretic mobility than the modest amounts of complex formed at low Mfd concentrations (figure 5.1 B compare lanes 5 and 7). A 32-mer would be expected to be long enough only to accommodate one or maybe two molecules of Mfd so it seems unlikely that this difference in electrophoretic mobility is due to the presence of multiple Mfd molecules binding to a single DNA oligomer. However, the Mfd protein might cover only a very short region of DNA, protein molecules could bind opposite faces of the helix, or alternatively additional Mfd molecules could bind to an already formed Mfd-DNA complex by interaction with bound Mfd protein rather than through contact with the DNA.

DNaseI footprinting of wild-type Mfd protein on a 250 bp DNA fragment

As one might expect of a protein that can move along DNA, Mfd binds DNA regardless of its sequence, presumably by contacting the DNA backbone. Despite Mfd binding to DNA in a sequence-independent manner, when DNA fragments were incubated with the C-terminally deleted Mfd protein, Mfd_{trunc}, regions of protection to DNaseI digestion were observed in addition to the appearance of hypersensitive sites (Selby and Sancar, 1995a). This digestion pattern is associated with the wrapping of DNA around a protein or with the binding of multiple protein molecules along the DNA fragment. The protected regions were located in the same positions on fragments that varied in length but possessed a common 3′ terminus, suggesting phasing of Mfd binding induced by a slight sequence preference or a structural element.

DNaseI footprinting was performed to investigate this observation further, and to determine whether the complexes that were observed in the presence of higher

concentrations of wild-type Mfd in DNA-binding EMSAs (see figure 5.1 A) were due to the binding of multiple Mfd molecules to the 250 bp fragment. The 250 bp *Eco*RI-*Bam*HI fragment of pSRc33 that was used in the DNA-binding EMSAs was also used for footprinting. In initial experiments the 5′(*Bam*HI) end of the fragment was labelled to observe the protection pattern on the template strand (with respect to the *lac*UV5 promoter present on the fragment). The end-labelled fragment was incubated with different concentrations of wild-type Mfd in the presence of ATPγS before treatment with DNaseI (figure 5.2 A).

In the presence of Mfd protein, the DNA was protected from digestion by DNaseI in two main regions centred approximately 130 and 235 bases from the *Bam*HI terminus. This protection was most apparent in the presence of 500 nM Mfd (compare lanes 1 and 4). As well as resulting in protection of two regions of the fragment, addition of Mfd created a number of hypersensitive sites. Hypersensitive sites are associated with bending or kinking of DNA by a protein, making certain positions of the DNA backbone more accessible to DNaseI. The presence of protected regions of DNA along with the appearance of hypersensitive sites is consistent with bending or wrapping of DNA by Mfd, or alternatively with the binding of more than one Mfd molecule to the 250 bp fragment. The binding of two molecules of Mfd to this fragment is consistent with a decrease in electrophoretic mobility in EMSAs of the Mfd-DNA complex formed at higher concentrations of Mfd than the mobility of the complex formed at lower Mfd concentrations (see figure 5.1).

Protection of the non-template strand was examined by labelling the 5′(*Eco*RI) end of the fragment. The footprinting patterns of the proteins containing substitutions

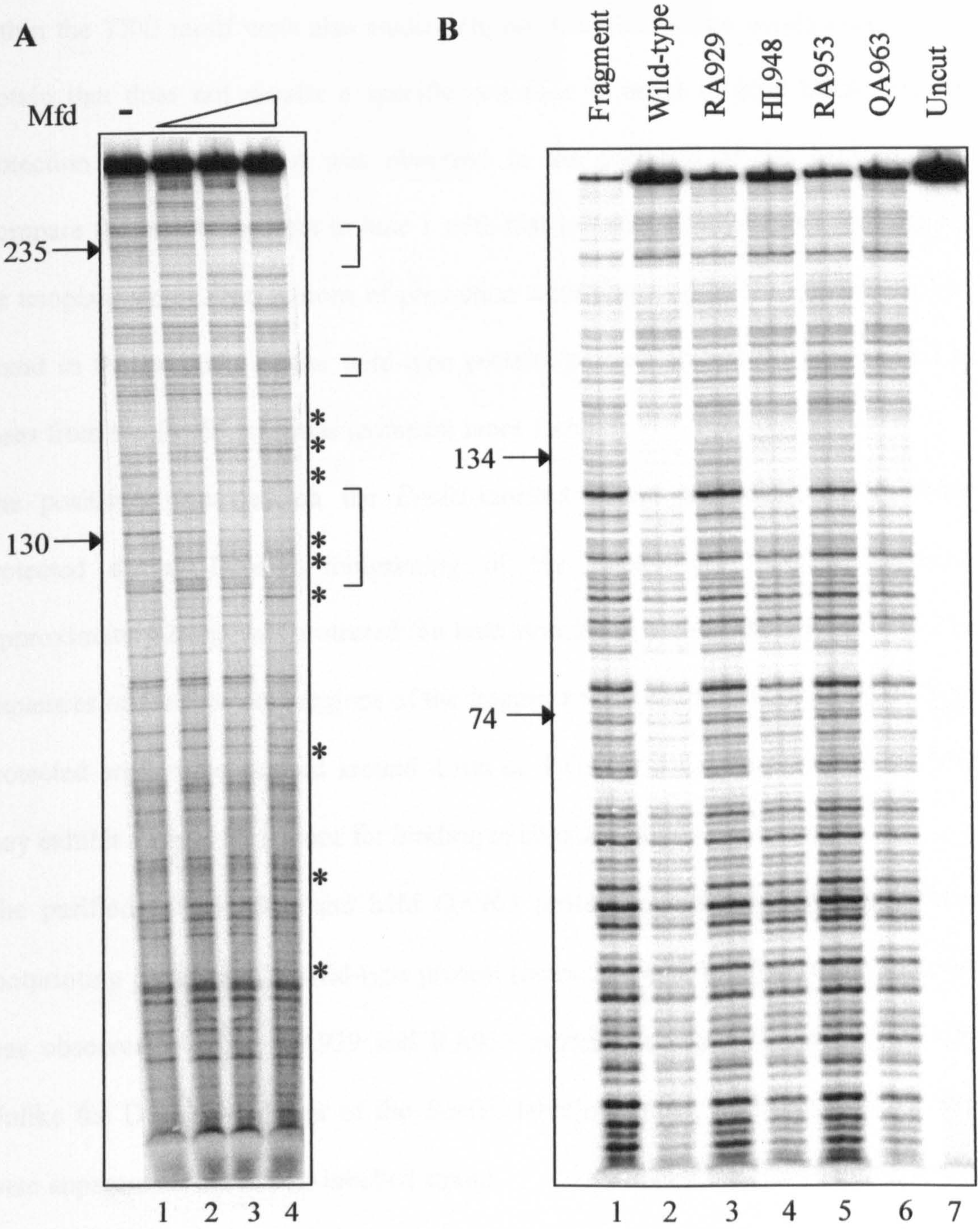


Figure 5.2. DNaseI footprints of the Mfd protein binding to pSRc33 fragment.

Mfd was bound to the *Eco*RI-*Bam*HI fragment of pSRc33 in the presence of ATPγS. The complexes were treated with DNaseI and the digestion pattern was analysed using a 6% denaturing TBE gel. (A) The 5′(*Bam*HI) end of the fragment was labelled and Mfd was added at a concentration of 50, 200 and 500 nM in lanes 2, 3 and 4 respectively. (B) The 5′(*Eco*RI) end of the fragment was labelled and was incubated with 500 nM of each purified protein. Lane 7 contains the fragment untreated by DNaseI. Representative gels are shown. Asterisks indicate hypersensitive sites and brackets indicate protected regions.

within the TRG motif were also studied (figure 5.2 B). As one would expect for a protein that does not require a specific sequence in order to bind DNA, general protection of the fragment was observed in the presence of the Mfd proteins (compare the intact fragment in lane 1 with that in lanes 2-6). As was the case for the template strand, two regions of protection were observed on the non-transcribed strand in the presence of the wild-type protein, centred around 74 bases and 134 bases from the *Eco*RI terminus (compare lanes 1 and 2).

The positions protected on the *Eco*RI-labelled strand are equivalent to those protected during DNaseI footprinting of the *Bam*HI-end labelled fragment. Approximately 25 bp was protected (on both strands) in each of these positions. The sequences of the protected regions of the fragment were examined (figure 5.3). Both protected areas were centred around a run of 4 Gs (or 4 Cs), suggesting that Mfd may exhibit a slight preference for binding to sites containing this sequence.

The purified Mfd HL948 and Mfd QA963 proteins produced the same DNaseI footprinting pattern as the wild-type protein (lanes 2, 4 and 6), however no footprint was observed with the RA929 and RA953 proteins (compare lanes 1, 3 and 5). Unlike the DNaseI footprint of the *Bam*HI-labelled strand, no hypersensitive sites were apparent on the *Eco*RI-labelled strand.

Taken together, the results of *in vitro* studies indicate the following DNA-binding activity of Mfd. Firstly it appears that multiple Mfd proteins are able to bind to a 250 bp DNA fragment. Footprinting studies on these complexes support the possibility of wrapping or bending of the DNA by the Mfd protein. The fact that Mfd is also able to bind to a 32-mer is consistent with the requirement for only 26 bp of DNA upstream of a transcription complex for its Mfd-mediated displacement and

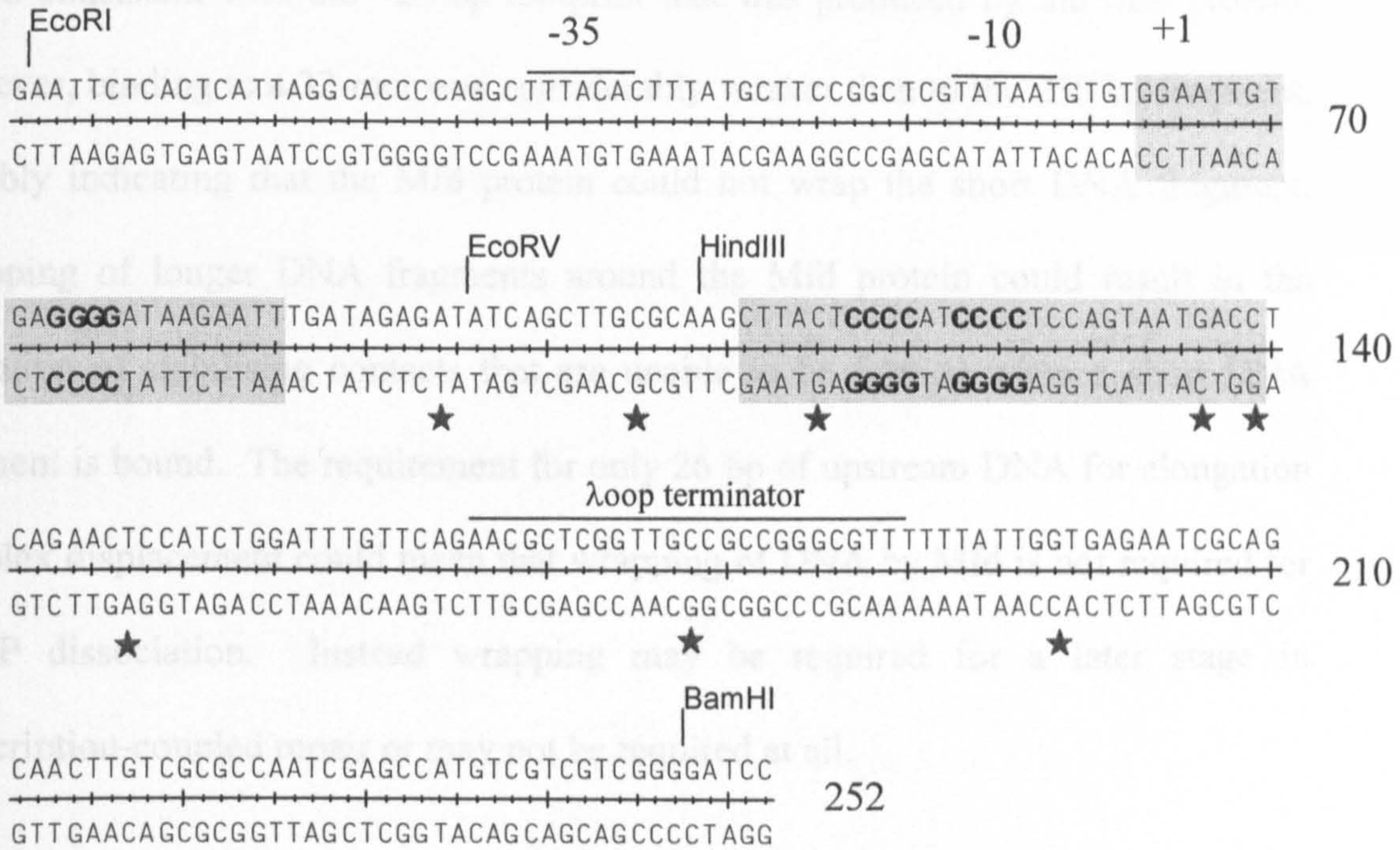


Figure 5.3. Location of DNaseI protection and hypersensitive sites.

The sequence of the *Eco*RI-*Bam*HI fragment of pSRc33 used for EMSAs and DNaseI footprinting is shown. The two major regions of protection in the presence of Mfd are shaded and stars indicate hypersensitive sites. The hypersensitive sites were located on the lower strand as shown. The runs of 4 Cs located in the two protected regions are shown in bold.

is also consistent with the ~25 bp footprint that was produced by the Mfd protein. However, binding to a 32-mer was considerably weaker than to the 250 bp fragment, possibly indicating that the Mfd protein could not wrap the short DNA fragment. Wrapping of longer DNA fragments around the Mfd protein could result in the formation of stabilising contacts that are unable to be formed when a short DNA fragment is bound. The requirement for only 26 bp of upstream DNA for elongation complex displacement could mean that wrapping of DNA by Mfd is not required for RNAP dissociation. Instead wrapping may be required for a later stage in transcription-coupled repair or may not be required at all.

Nucleotide binding and hydrolysis by the Mfd protein

The seven conserved helicase motifs of Mfd classify it as a member of helicase superfamily 2 (Gorbalenya and Koonin, 1993). Helicase motifs are responsible for binding and hydrolysis of ATP and it has been demonstrated that Mfd is able to hydrolyse both ATP and dATP in order to displace a stalled elongation complex (Selby and Sancar, 1993a).

Nucleotide specificity of the Mfd protein

Nucleotide hydrolysis is required for elongation complex displacement by Mfd. Although it is known that Mfd can utilise ATP (or with a decreased efficiency, dATP) to power elongation complex displacement, it is not known whether Mfd can utilise CTP, GTP or UTP.

The displacement of elongation complexes in the presence of an excess of different nucleotides was analysed. The *Eco*RI-*Bam*HI fragment of pSRca19 contains the *lac*UV5 promoter from which a 19 nucleotide transcript can be produced by the addition of only the initiating dinucleotide ApU, and UTP and GTP (figure 5.4 A).

A

-35 -10 +1

.....CCCCAG**CTTTACA**CTTTATGCTTCCGGCTCG**TATAAT**GTGTGGA

+20

TTTGTGTGGGGTTGTGGTCTTGATAGAGATATC.....

B

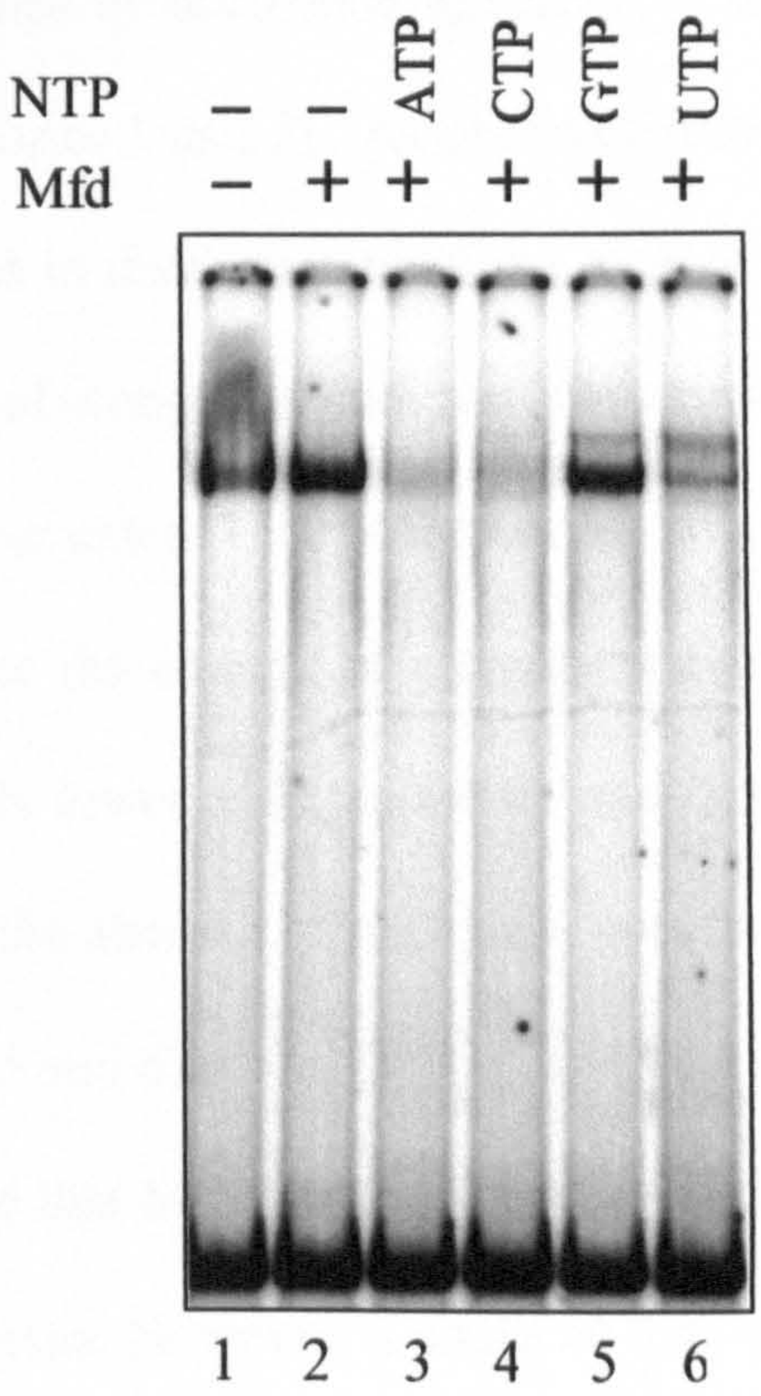


Figure 5.4. Mfd-mediated displacement of elongation complexes stalled on the *Eco*RI-*Bam*HI fragment of pSRca19 in the presence of different nucleotides.

(A) Part of the *Eco*RI-*Bam*HI fragment of pSRca19. Elongation complexes will stall by nucleotide starvation at +20 in the presence of the initiating nucleotide ApU, UTP and GTP. (B) An EMSA analysing elongation complexes stalled by nucleotide starvation (on the above fragment) treated with 250 nM Mfd protein and 2 mM nucleotide for 15 minutes at 37°C.

Transcription stalls at +19 on this fragment in the absence of CTP due to nucleotide starvation. In the presence of CTP, but in the absence of ATP, transcription of a further four bases is possible before transcription will stall at +23 due to ATP starvation. Transcription complexes were formed at +19 in the absence of ATP and CTP and Mfd protein was added with 2 mM of the nucleotide to be examined. The complexes were analysed by EMSA (figure 5.4 B).

Addition of Mfd in the absence of nucleotide appeared to stabilise the elongation complex in the gel (compare lanes 1 and 2). As observed previously, the addition of 2 mM ATP with Mfd resulted in displacement of the elongation complex (compare lanes 2 and 3). The amount of elongation complex present was also reduced in the presence of CTP and to a lesser extent UTP (compare lanes 2, 4 and 6). Addition of GTP with Mfd did not reduce the amount of elongation complex present (lane 5). An additional complex, with lower electrophoretic mobility than the elongation complex that was formed in the absence of nucleotide was formed when CTP, UTP or GTP were added (lanes 4, 5 and 6).

From this analysis it appeared that Mfd could utilise CTP and to a lesser extent UTP to displace elongation complexes. However, it could be that addition of CTP or UTP destabilised the elongation complex in an Mfd-independent manner.

To examine this possibility, elongation complexes, labelled by incorporation of [^{32}P] αUTP into the transcript, were formed on the *Eco*RI-*Bam*HI fragment of pSRca19, so that only elongation complexes were visualised. The complexes were stalled by nucleotide starvation at +19 and incubated in the absence of nucleotide or with 2 mM of each nucleotide in the absence and presence of Mfd. The resulting complexes were analysed by EMSA (figure 5.5 A). The effect of nucleotide addition

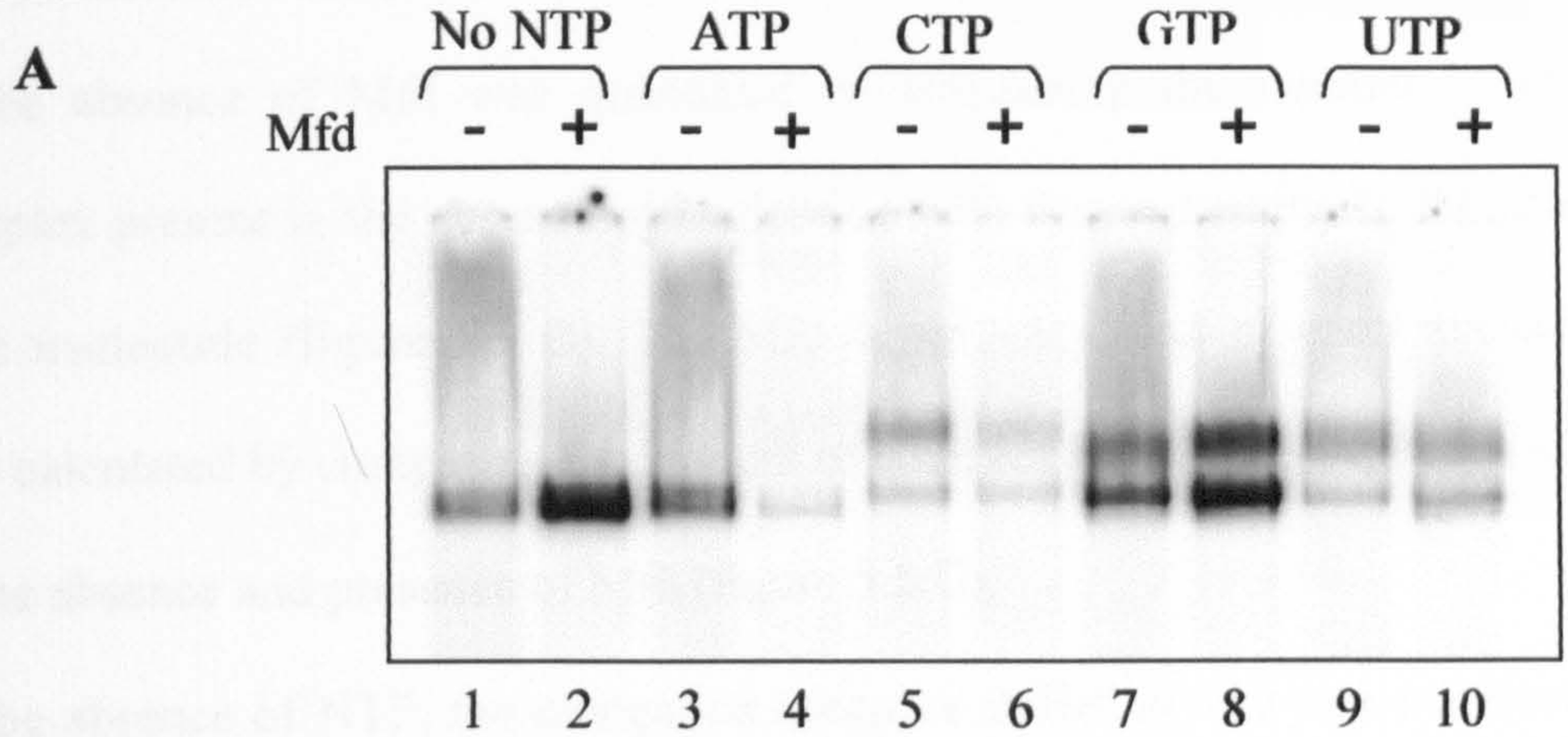


Figure 5.5. Nucleotide utilisation of Mfd during elongation complex displacement.

(A) An EMSA of elongation complex displacement. Stalled elongation complexes with labelled UTP incorporated into the transcript were formed on the *Eco*RI-*Bam*HI fragment of pSRca19. Where indicated 2 mM nucleotide and 250 nM Mfd was added to the stalled complexes, incubated for 15 minutes at 37°C and analysed by EMSA using a 4.5% acrylamide TBE gel. (B) Mfd-independent effect of addition of 2 mM nucleotide to stalled elongation complexes. The amount of complex is expressed as a percentage of the elongation complex present in the absence of nucleotide. Both complexes were taken into account. (C) Quantification of the effect of Mfd in the presence of each nucleotide. Complexes are expressed as a percentage of the elongation complex in the presence of each nucleotide in the absence of Mfd.

in the absence of Mfd was quantified by comparing the amount of elongation complex present in the absence of nucleotide with that remaining in the presence of each nucleotide (figure 5.5 B). The Mfd-dependent effect of each nucleotide was also calculated by comparing the amount of complex remaining with each nucleotide in the absence and presence of Mfd (figure 5.5 C).

In the absence of NTP, the elongation complex stalled at +19 was stabilised within the gel by the presence of 250 nM Mfd protein (compare lanes 1 and 2). The stabilising effect of Mfd did not require the presence of nucleotide and therefore occurs under conditions in which the Mfd protein is unable to form a stable complex with DNA. The elongation complex remained stable in the presence of ATP but was displaced by the addition of Mfd protein with the ATP (lanes 3 and 4). Addition of CTP or UTP alone to the elongation complex resulted in its destabilisation (compare lanes 1, 5 and 9 and see figure 5.5 B).

With the exception of ATP, addition of nucleotide to the elongation complex resulted in the formation of a second complex that possessed lower electrophoretic mobility (lanes 5, 7 and 9). Since the RNA is labelled, this additional complex must be a functional elongation complex that has either changed conformation or has moved along the DNA, for example by backtracking or reading-through the stall site by misincorporation into the transcript due to the high concentration of nucleotide that was added. It is unclear why addition of CTP, GTP and UTP but not ATP resulted in the formation of this second complex.

In order to consider the ability of Mfd to utilise each nucleotide for elongation complex displacement, the destabilisation of the elongation complex by CTP and UTP was taken into account i.e. the amount of elongation complex present following nucleotide addition was compared to the amount remaining following the addition of

nucleotide and Mfd. No significant decrease in elongation complex was observed upon Mfd addition in the presence of GTP or UTP (figure 5.5 C compare lanes 7 and 8 plus 9 and 10), however the intensity of the elongation complex band in the presence of CTP decreased upon Mfd addition (compare lanes 5 and 6). Upon quantification it was seen that approximately 50% of elongation complexes were displaced by Mfd in the presence of CTP after 15 minutes (figure 5.5 C). This indicated that Mfd is able to utilise CTP to cause RNAP dissociation.

CTP usage was less efficient than ATP since Mfd was able to displace approximately 90% of stalled complexes within 5 minutes in the presence of ATP (figure 4.12). It may be relevant that the polymerase had been stalled by CTP starvation and the destabilisation that was observed upon CTP addition could therefore be a consequence of the addition of the next nucleotide at a high concentration causing incorporation of the NTP into the growing transcript. Since addition of CTP to elongation complexes stalled by CTP starvation in itself caused some dissociation of elongation complexes, the ability of Mfd to use CTP could be confirmed using a template where transcription was stalled by the omission of either ATP, GTP or UTP or by walking the RNA polymerase to a stall site.

RNA polymerase is unable to incorporate dNTPs into the transcript (Svetlov *et al.*, 2004) and as such any destabilising effects induced by the forward motion of the stalled elongation complex in response to addition of high concentrations of NTPs can be avoided by the use of dNTPs to test the nucleotide base specificity of Mfd.

Preliminary experiments examining the ability of Mfd to utilise dNTPs to bring about elongation complex displacement showed that the stability of the elongation complex was not affected by the addition of any of the dNTPs. The data also suggested that dATP and dCTP could be utilised by the Mfd protein to dissociate

RNAP (data not shown). Mfd-mediated displacement of elongation complexes has been shown to occur in the presence of dATP (Park *et al.*, 2002) and if Mfd is able to use both ATP and less efficiently CTP (see figure 5.5) it is not surprising that dCTP is able to be hydrolysed by Mfd. The utilisation of dATP and dCTP could be confirmed by further experiments incorporating label into the nascent RNA to view only elongation complexes.

In summary wild-type Mfd appears to be able to use ATP and dATP and to a lesser extent CTP and dCTP to bring about displacement of stalled elongation complexes.

The Q-tip motif

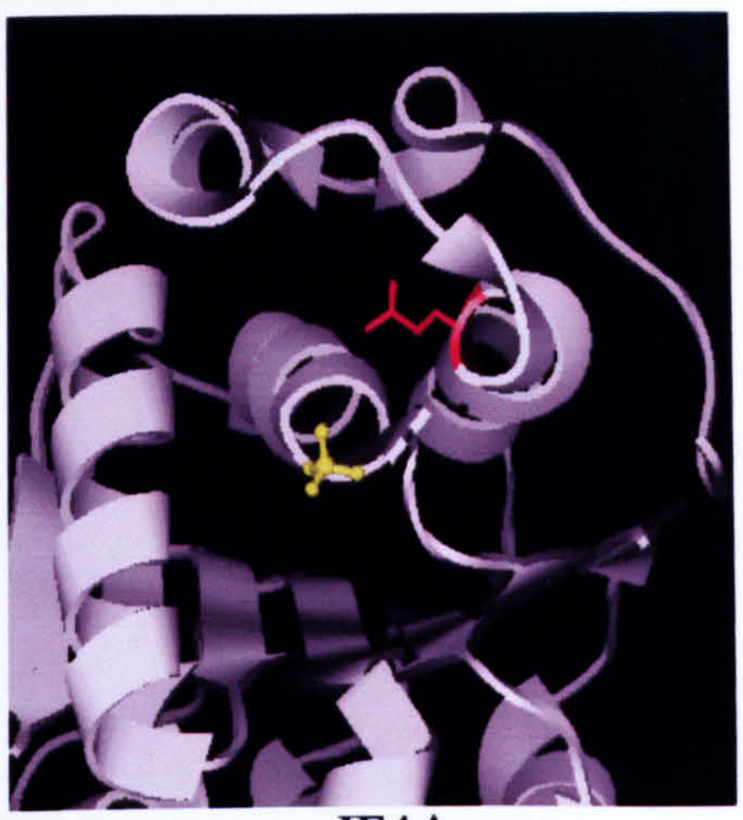
Recently an extra helicase motif has been identified upstream of the traditional helicase motifs (Tanner *et al.*, 2003). This motif, termed the Q-motif, was discovered in the DEAD-box RNA helicase family, which act on short double-stranded regions of RNA to ensure correct folding of RNA molecules. Within the DEAD-box helicase family the Q-motif consists of a group of nine conserved amino acids including an invariant glutamine (Tanner *et al.*, 2003). In addition a highly conserved aromatic amino acid is located upstream of the Q-motif. Substitution of these conserved amino acids in the essential yeast protein, eIF4A, was lethal *in vivo* and the mutated protein was defective in ATP binding *in vitro.* Mutation of the Q-motif in a further two DEAD-box helicases, Ded1 and the flavivirus NS3 helicase, confirmed that it was essential for normal ATP binding and hydrolysis (Cordin *et al.*, 2004; Gallivan and McGarvey, 2003). In the crystal structure of eIF4A the glutamine residue of the Q-motif is located in the binding pocket for the nucleotide base of ATP (Caruthers *et al.*, 2000). Based on the fact that substitution of the conserved glutamine in some essential yeast RNA and DNA helicases was not lethal and that the NS3 RNA helicase from HCV, which does not possess this glutamine,

shows non-specific NTP usage, it was proposed that the motif may play a role in nucleotide affinity or specificity.

Although the Q-motif is unique to the DEAD-box helicase family, features of this motif can be found in other helicases. Upon examination of the crystal structures of UvrB, RecG, PcrA, eIF4A and RecQ, a glutamine residue was located within hydrogen-bonding distance of the nucleotide base (figure 5.6) (Dr. M. Szczelkun, personal communication). In addition this glutamine residue is located in the same structural context in each protein i.e. at the N-terminus of an α-helix. Sequence alignments of a number of SF1, SF2 and AAA+ proteins showed the presence of a conserved glutamine a variable distance upstream of motif I. In addition a small amino acid (alanine, serine or valine) was located three amino acids downstream (figure 5.7). The structurally conserved glutamine and small downstream amino acid were collectively termed the Q-tip (Dr. M. Szczelkun, personal communication).

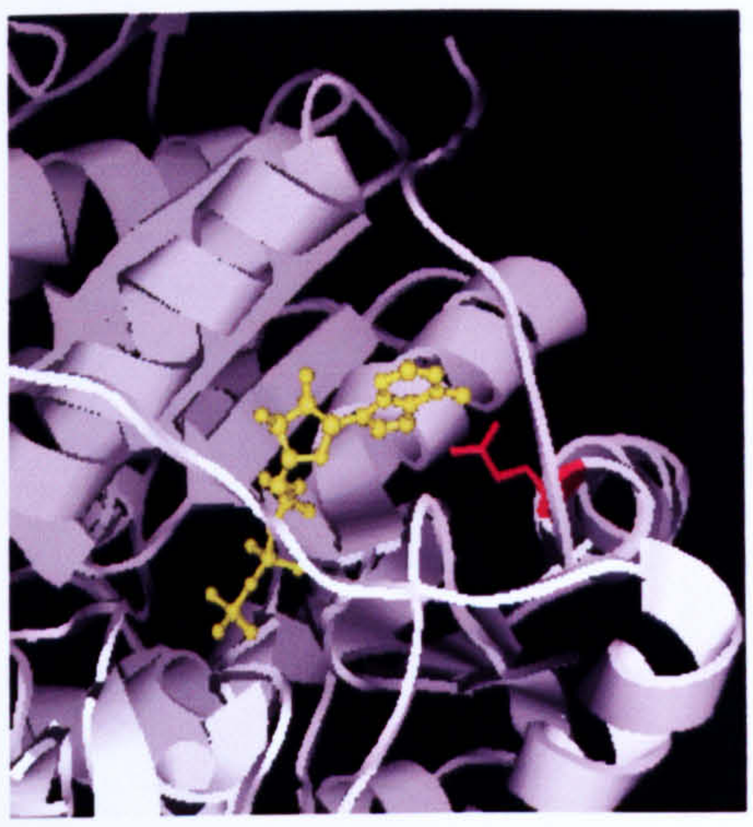
A glutamine located 22 amino acids upstream of motif I is conserved in all Mfd sequences examined. In a structural model of the helicase domains, based on homology to the helicase domains of RecG, this glutamine is located at the end of an α-helix close to the base of the bound nucleotide (figure 5.6). A neighbouring aspartate residue is also completely conserved across Mfd proteins from different species. This aspartate is not present in RecG, a member of the same helicase subfamily as Mfd.

The Q-tip has been shown to be required for ATP binding and hydrolysis in members of the DEAD-box RNA helicase family but as yet its importance in other helicase subfamilies has not been established. Therefore the requirement of these two Q-tip residues for *in vivo* Mfd activity and the effect of their substitution on

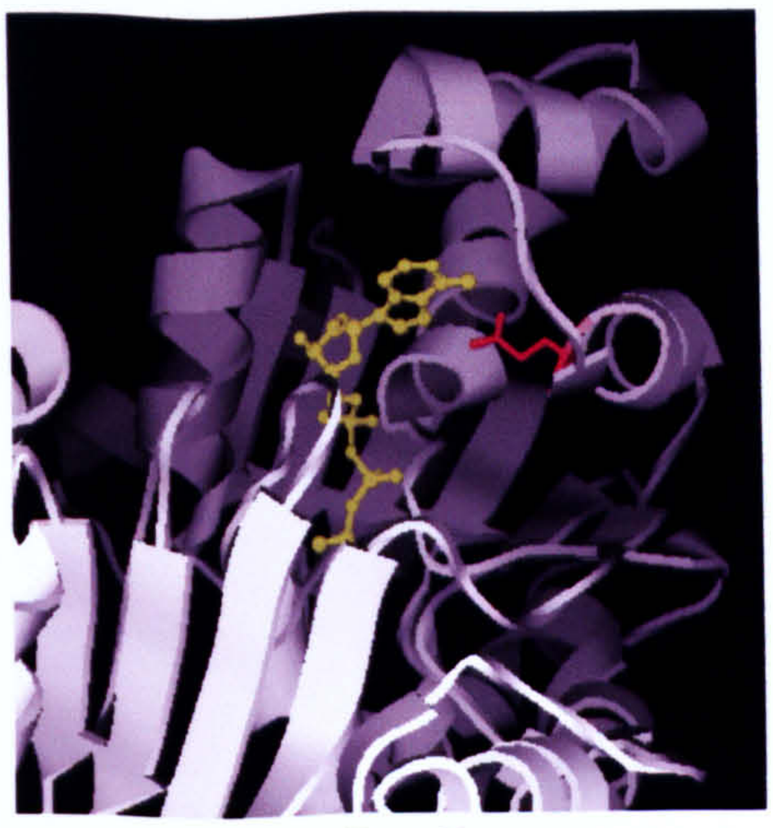


eIF4A

PcrA

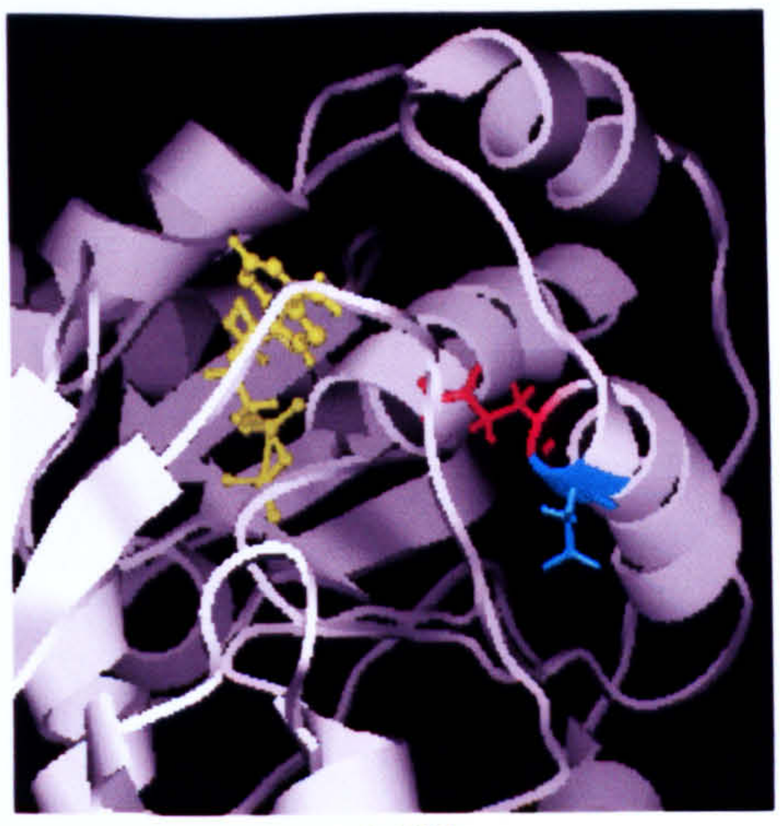


UvrB



RecQ

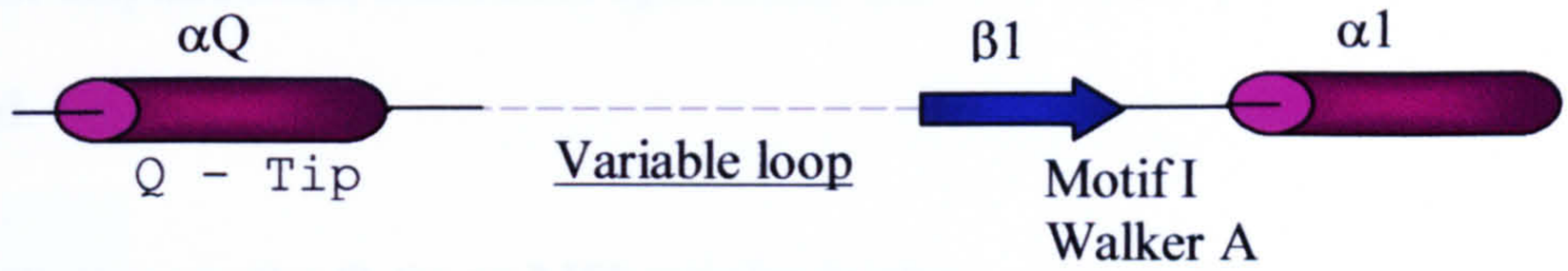
RecG

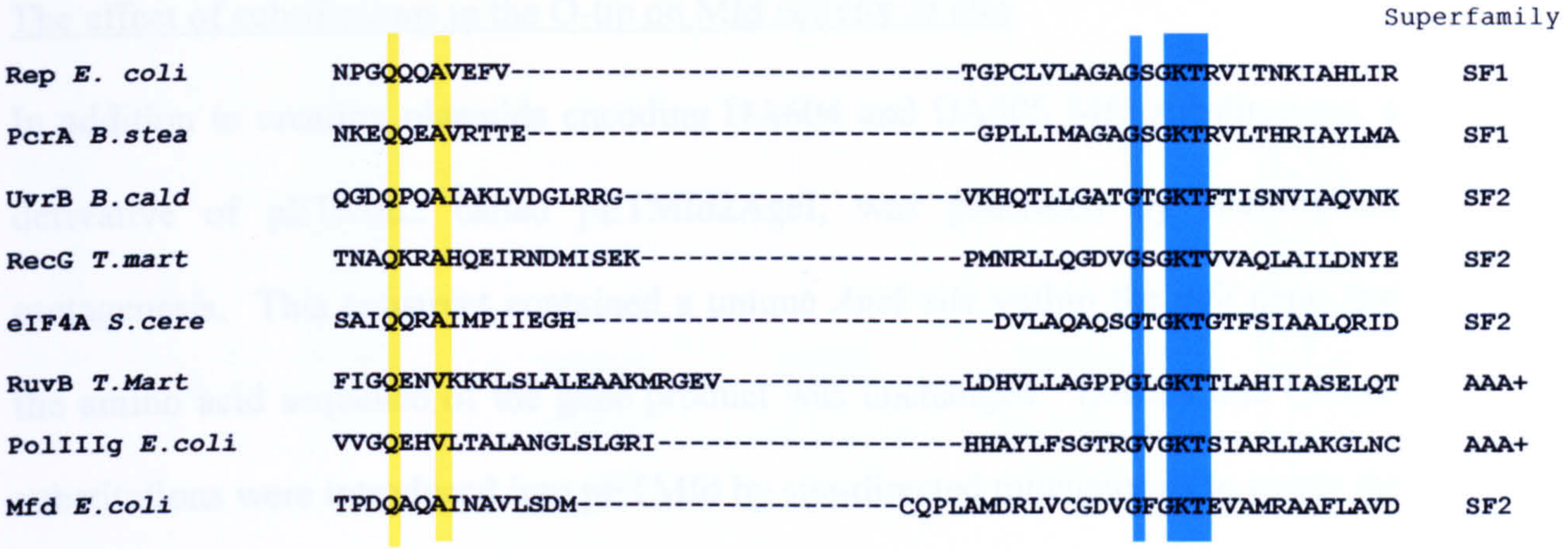


Mfd

Figure 5.6. The Q-tip

The structures of eIF4A (Caruthers *et al.*, 2000), PcrA (Velankar *et al.*, 1999), UvrB (Theis *et al.*, 1999), RecQ (Bernstein *et al.*, 2003), RecG (Singleton *et al.*, 2001) and the model of Mfd show that the conserved glutamine upstream of motif I is within hydrogen bonding distance of the nucleotide base and is located at the end of a helix. The nucleotide is shown in yellow, the glutamine in red and the conserved aspartate of Mfd in cyan.





S

MFD Family FxxxFPxETxDQxxAIxxxxxD

Figure 5.7. Sequence alignment of the Q-tip

A conserved glutamine residue is located at the N-terminus of an α-helix. The structural context of the glutamine is conserved but it is located a variable distance upstream of motif I. A small amino acid (alanine, serine or valine) is located three amino acids downstream of the glutamine. An aspartate residue adjacent to the conserved glutamine in Mfd sequences is 100% conserved.

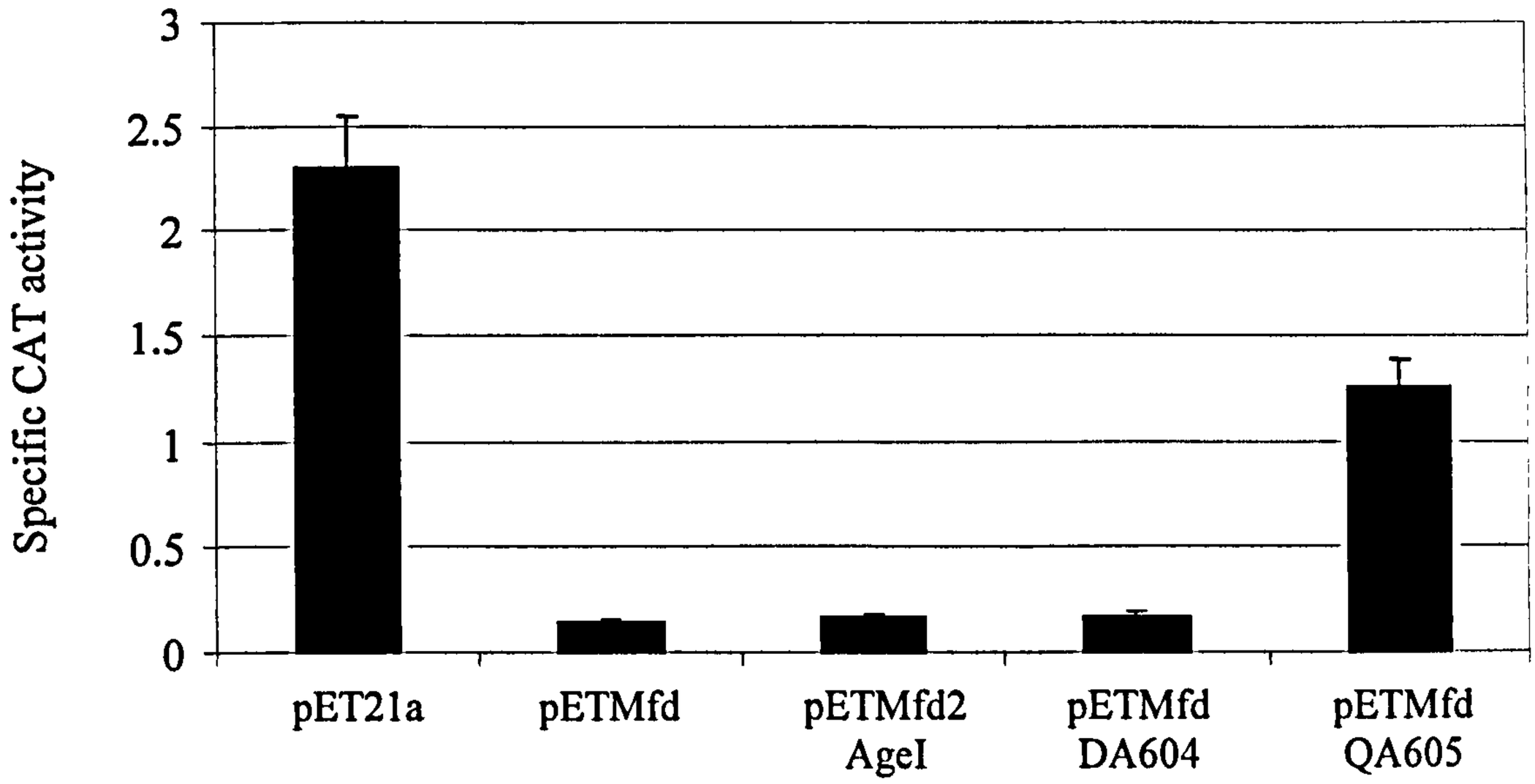
elongation complex displacement, nucleotide specificity and ATP hydrolysis *in vitro* has been examined.

The effect of substitutions in the Q-tip on Mfd activity *in vivo*

In addition to creating plasmids encoding DA604 and QA605 Mfd substitutions, a derivative of pETMfd2 called pETMfd2AgeI, was generated by site-directed mutagenesis. This construct contained a unique *Age*I site within the *mfd* gene, but the amino acid sequence of the gene product was unchanged. DA604 and QA605 substitutions were introduced into pETMfd by site-directed mutagenesis to create the plasmids pETMfdDA604 and pETMfdQA605 respectively.

The effect of substitutions in the Q-tip on Mfd function *in vivo* was tested in the *cat* roadblock repression assay. UNCNOMFD cells containing the reporter construct pRCB-CAT were transformed with the empty vector pET21a, pETMfd, pETMfdAgeI, pETMfdDA604 and pETMfdQA605. The cells were grown in minimal media containing the appropriate antibiotics until mid-exponential phase when the cells were harvested and their CAT activity measured (figure 5.8).

Silent mutagenesis within codon 481 of the *mfd* gene to introduce an *Age*I site did not affect the ability of Mfd to increase the efficiency of roadblock repression. Upon transformation with pETMfd or pETMfd2AgeI, the CAT activity was 16.4 or 13.5-fold lower, respectively than cells that had been transformed with the empty vector. Transformation with pETMfdDA604 also resulted in a level of CAT activity equivalent to that seen with the wild-type *mfd* plasmid. The cells transformed with pETMfdQA605 possessed higher CAT activity than those expressing wild-type Mfd, but the CAT activity remained lower than in the complete absence of Mfd. pETMfdQA605 resulted in a 1.8-fold decrease in CAT activity in comparison to



Plasmid	CAT activity	Fold Mfd effect +Mfd/-Mfd
pETMfd	0.14+/-0.02	16.4
pETMfd2AgeI	0.17+/-0.01	13.5
pETMfdDA604	0.16+/-0.03	14.3
pETMfdQA605	1.25+/-0.14	1.84

Figure 5.8. The effect of substitutions within the Q-tip of Mfd on *in vivo* roadblock repression.

UNCNOMFD cells containing the roadblock reporter construct pRCB-CAT were transformed with the empty vector pET21a, pETMfd, pETMfd2AgeI, pETMfdDA604 or pETMfdQA605. The cells were grown in minimal media containing kanamycin, tetracycline and ampicillin until mid-exponential phase. Values shown are the average of three experiments and are shown with s.d. Units are nmol chloramphenicol acetylated/min/mg protein.

pET21a and approximately a 9-fold increase in CAT activity in comparison to pETMfd. These data suggest that substitution of the conserved Q-tip glutamine does not completely abolish Mfd function *in vivo*, but decreases the ability of the Mfd protein to increase the efficiency of roadblock repression. The simplest explanation of these data is that the ability of Mfd to displace elongation complexes has been reduced by substitution of Q605. Mfd DA604 was able to increase the efficiency of roadblock repression to the same extent as wild-type Mfd, indicating that this substitution does not abolish the ability of Mfd to displace elongation complexes.

The effect of substitutions in the Q-tip on Mfd activity *in vitro*

Although the defect in elongation complex displacement *in vivo* by Mfd QA605 is expected to be due to a defect in ATP hydrolysis, the process of RNAP displacement requires the Mfd protein to be proficient in RNAP binding, DNA binding, ATP binding and coupling of ATP hydrolysis to RNAP displacement and as such the defect in the QA605 Mfd protein could be in any of these activities. Therefore the Mfd DA604 and Mfd QA605 proteins were purified and their ability to displace elongation complexes, their nucleotide specificity, their ATP hydrolysis activity and their ability to bind DNA was tested *in vitro*.

Role of the Q-tip of Mfd in the nucleotide specificity of elongation complex displacement

The Q-tip is required for the helicase activity of DEAD-box helicases but the importance of this motif in the function of other helicases is undetermined. Mfd proteins containing substitutions in the Q-tip were analysed to test whether the Q-tip of Mfd is required for the displacement of elongation complexes *in vitro*. The Q-tip is located close to the nucleotide binding site and therefore substitutions in this motif

could result in altered nucleotide specificity. Therefore the nucleotide-dependence of elongation complex displacement by Mfd proteins containing substitutions within the Q-tip was tested to determine whether these proteins possessed the same nucleotide binding specificity as the wild-type protein.

The DA604 substitution did not affect elongation complex displacement by Mfd *in vivo*. However, previous studies have shown that the *in vivo* assay is insensitive to mutants that are partially defective in RNAP dissociation. To examine the effect of substitution of D604 *in vitro*, elongation complexes were stalled by nucleotide starvation on the *Eco*RI-*Bam*HI fragment of pSRca19. Nucleotides were added to these complexes along with Mfd DA604 protein and the resulting complexes were analysed by EMSA (figure 5.9). The pattern of elongation complex displacement was the same as for the wild-type protein (see figure 5.4 B). The presence of Mfd DA604 stabilised the elongation complex in the absence of nucleotide (lanes 1 and 2). Addition of Mfd with ATP, CTP and to a lesser extent UTP resulted in a decrease in the amount of elongation complex present (compare lanes 1, 3, 4 and 6). In the case of the wild-type protein the decrease in elongation complex in the presence of UTP was shown to be due solely to the presence of the nucleotide, while the decrease upon addition of CTP and Mfd was shown to be partially the result of nucleotide addition and partially an Mfd effect. This seems likely to be the case for the DA604 protein also but could be confirmed using complexes containing labelled RNA and testing the effect of each nucleotide in the absence and presence of Mfd. No decrease in the amount of elongation complex was observed in the presence of GTP (lane 5). As such the DA604 Mfd protein appears to behave like the wild-type protein in displacement of elongation complexes both *in vivo* and *in vitro*.

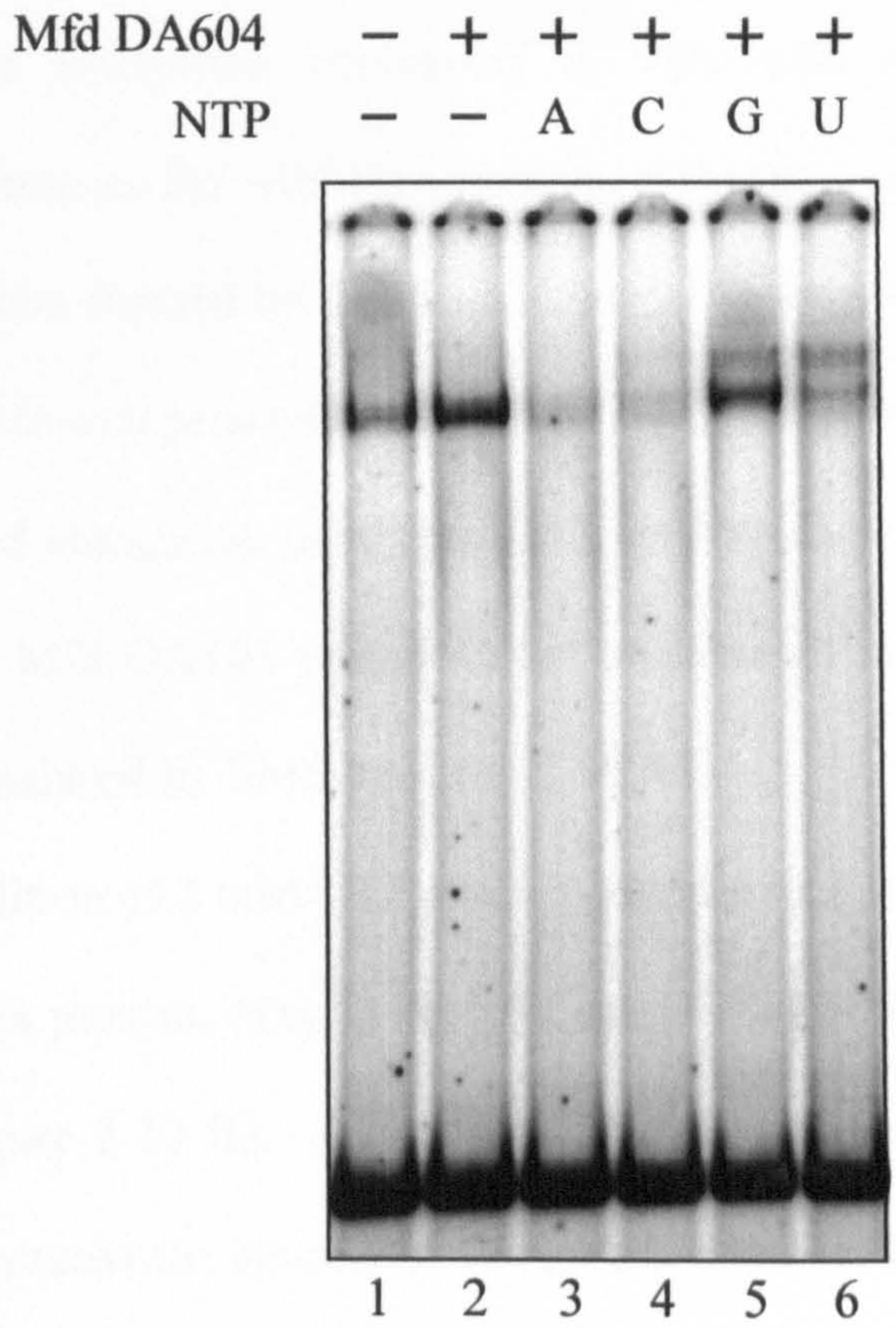


Figure 5.9. Nucleotide usage of Mfd DA604 in elongation complex displacement.

Elongation complexes were stalled by nucleotide starvation on the *Eco*RI-*Bam*HI fragment of pSRca19. Where indicated 2 mM nucleotide and 250 nM purified DA604 Mfd protein was added to the stalled complexes. Reactions were incubated for 15 minutes at 37°C following Mfd addition before being analysed by EMSA on a 4.5% acrylamide/TBE gel.

Mfd QA605 was defective in elongation complex formation *in vivo*. To examine its ability to displace elongation complexes *in vitro* and whether its nucleotide specificity is the same as the wild-type protein, elongation complexes with labelled RNA transcripts were formed on the *Eco*RI-*Bam*HI fragment of pSRca19. In order to determine the Mfd-independent effect of nucleotide addition, nucleotide alone was added to the stalled elongation complexes and the Mfd-dependent effect was tested by addition of the Mfd QA605 protein in the presence of nucleotide. The resulting complexes were analysed by EMSA (figure 5.10 A).

As previously, addition of 2 mM CTP or UTP resulted in a decrease in the amount of elongation complex present, even in the absence of Mfd protein (compare lanes 1, 5 and 9 and see figure 5.10 B). Addition of Mfd to complexes with GTP or UTP present, did not decrease the amount of elongation complex that was observed (lanes 7, 8, 9 and 10); this was also true for the wild-type protein. However, in contrast to the wild-type and Mfd DA604 proteins, following addition of ATP or CTP there was no reduction in the amount of elongation complex upon incubation with Mfd QA605 (lanes 3, 4, 5 and 6 and figure 5.10 C). The Mfd QA605 protein was unable to displace stalled elongation complexes *in vitro* using any of the nucleotides; this Q-tip substitution abolishes the function of Mfd rather that simply altering its nucleotide specificity.

Preliminary data examining the ability of Mfd QA605 to utilise dNTPs to power elongation complex displacement indicated that elongation complexes were not displaced by the addition of Mfd QA605 in the absence of nucleotide or in the presence of any of the four dNTPs. Substitution of Q605 appears to have abolished the ability of the Mfd protein to utilise dATP or dCTP as well as ATP or CTP for

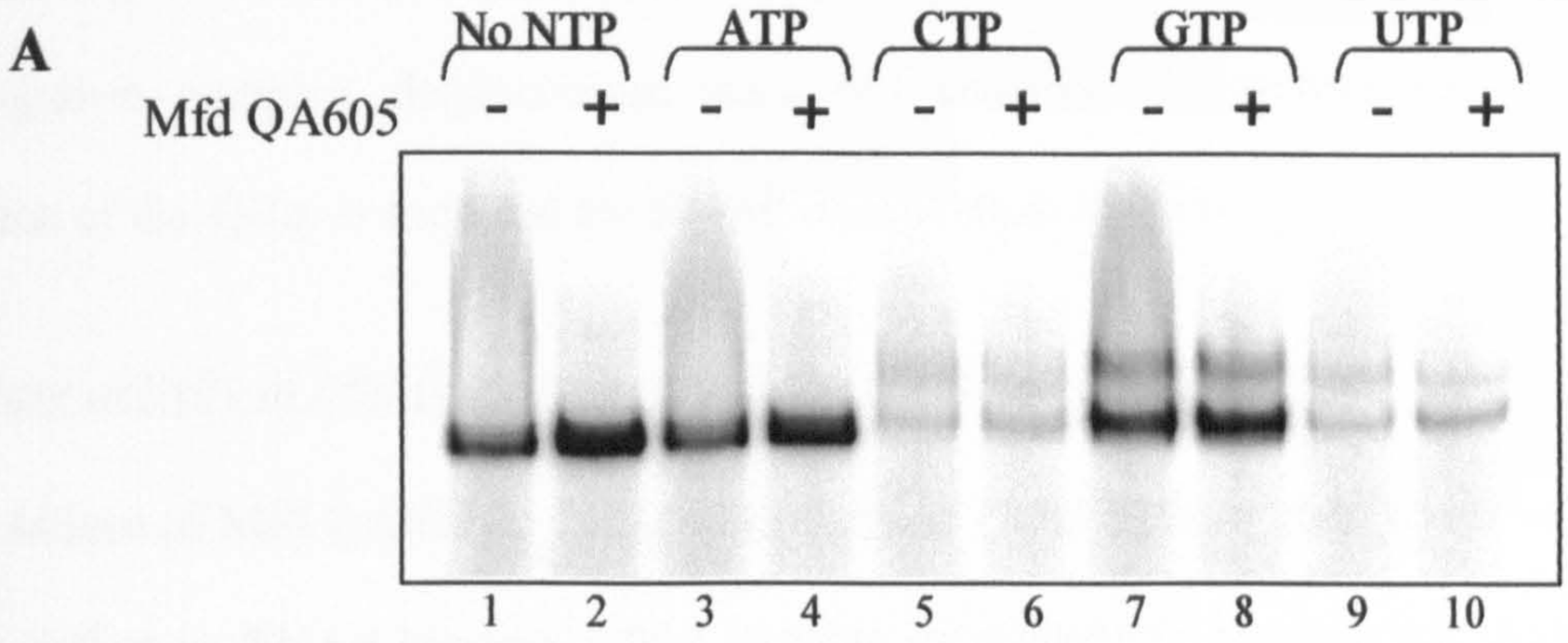


Figure 5.10. Nucleotide specificity of elongation complex displacement by Mfd QA605.

(A) An EMSA of elongation complex displacement. Stalled elongation complexes with labelled UTP incorporated into the transcript were formed on the *Eco*RI-*Bam*HI fragment of pSRca19. Where indicated 2 mM nucleotide and 250 nM Mfd was added to the stalled complexes, incubated for 15 minutes at 37°C and analysed by EMSA using a 4.5% acrylamide/TBE gel. (B) Mfd-independent effect of addition of 2 mM nucleotide to stalled elongation complexes. Complexes are expressed as a percentage of the elongation complex present in the absence of nucleotide. Both complexes were taken into account. (C) Quantification of the effect of Mfd in the presence of each nucleotide. Complexes are expressed as a percentage of the elongation complex in the presence of each nucleotide in the absence of Mfd.

elongation complex displacement (data not shown). Therefore the glutamine residue of the Q-tip is required for RNAP dissociation *in vitro*.

ATPase activity of Mfd proteins containing substitutions in the Q-tip

The failure of Mfd QA605 to displace elongation complexes *in vitro* could be caused by a defect in RNAP binding, DNA binding or nucleotide binding and hydrolysis. The Q-tip residues are located near the nucleotide-binding site of Mfd, therefore in order to determine whether substitution of D604 and Q605 affect ATP hydrolysis by Mfd and therefore whether the failure of Mfd QA605 to displace elongation complexes is due to a defect in nucleotide hydrolysis, the ATPase activity of the purified proteins containing substitutions within the Q-tip was assayed (figure 5.11). While the wild-type protein hydrolysed ATP with a k_{cat} of approximately 20 min^{-1} (figure 5.11 pink squares), the rate of ATP hydrolysis by both DA604 and QA605 Mfd proteins was not significantly above the background level.

This indicates that both of these substitutions reduce the ability of Mfd to hydrolyse ATP. This could be due to a defect in ATP binding, ATP hydrolysis or product release; alternatively it could be caused by the substitutions disrupting the fold or stability of the protein. Both of the substitutions resulted in proteins that were still at least partially able to displace elongation complexes *in vivo*. No difference was observed between Mfd DA604 and the wild-type protein in *in vivo* roadblock repression assays. This may indicate that substitution of D604 results not in the complete abolition of ATPase activity, but in a decrease in ATPase activity to below the level of detection of the assay that was used. The inability of Mfd QA605 to displace stalled elongation complexes *in vitro* and its decreased ability to increase the efficiency of roadblock repression *in vivo* is most likely to be caused by its defect

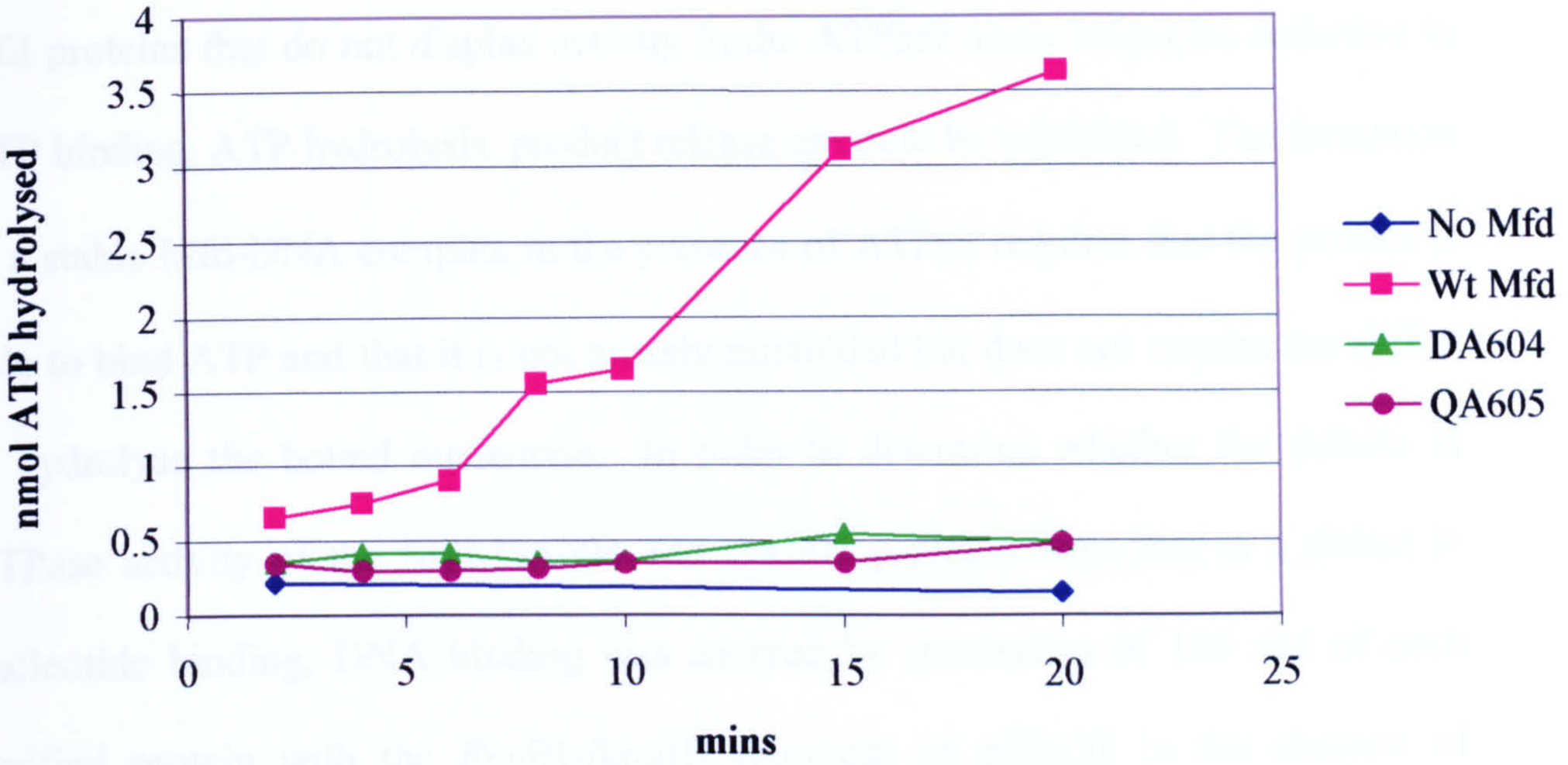


Figure 5.11. ATPase activity of the Mfd proteins containing substitutions in the Q-tip.

Assays contained 1 μM Mfd protein, 2 mM ATP spiked with 1 μCi [γ-^{32}P] ATP in a 10 μl reaction volume. The reactions were incubated at 37°C and were stopped at the time points indicated by the addition of 30 mg/ml charcoal in 40 mM HCl. Unhydrolysed ATP adsorbed to the activated charcoal and the hydrolysed ATP was quantified by Cerenkov counting. Assays were performed in duplicate and the averages are shown.

in ATP hydrolysis.

DNA binding activity of Mfd proteins containing substitutions in the Q-tip

Mfd proteins that do not display activity in the ATPase assay might be defective in ATP binding, ATP hydrolysis, product release or could be misfolded. The formation of a stable Mfd-DNA complex in the presence of ATPγS requires that the protein is able to bind ATP and that it is not grossly misfolded but does not require the ability to hydrolyse the bound nucleotide. In order to determine whether the defects in ATPase activity of the Mfd DA604 and QA605 proteins were due to a defect in nucleotide binding, DNA binding was assayed by incubation of 100 nM of each purified protein with the *Eco*RI-*Bam*HI fragment of pSRc33 in the absence of nucleotide or the presence of ATP or ATPγS. The resulting complexes were analysed by EMSA (figure 5.12 A). As observed previously, the wild-type protein only formed a stable complex with DNA in the presence of ATPγS (compare lanes 1-3) and this complex had a lower electrophoretic mobility than the complex that was formed upon incubation of Mfd RA953 (TRG motif substitution) with ATP (compare lanes 3 and 4). Mfd DA604 did not form a stable complex with DNA in the absence of nucleotide or in the presence of ATP (lanes 5 and 6). An Mfd DA604-DNA complex was observed in the presence of ATPγS and this complex possessed the same electrophoretic mobility as the complex formed with 100 nM Mfd RA953 in the presence of ATP (compare lanes 4 and 7). Mfd QA605 did not form a stable complex with DNA in the absence of nucleotide or in the presence of 2 mM ATP or ATPγS (lanes 8-10).

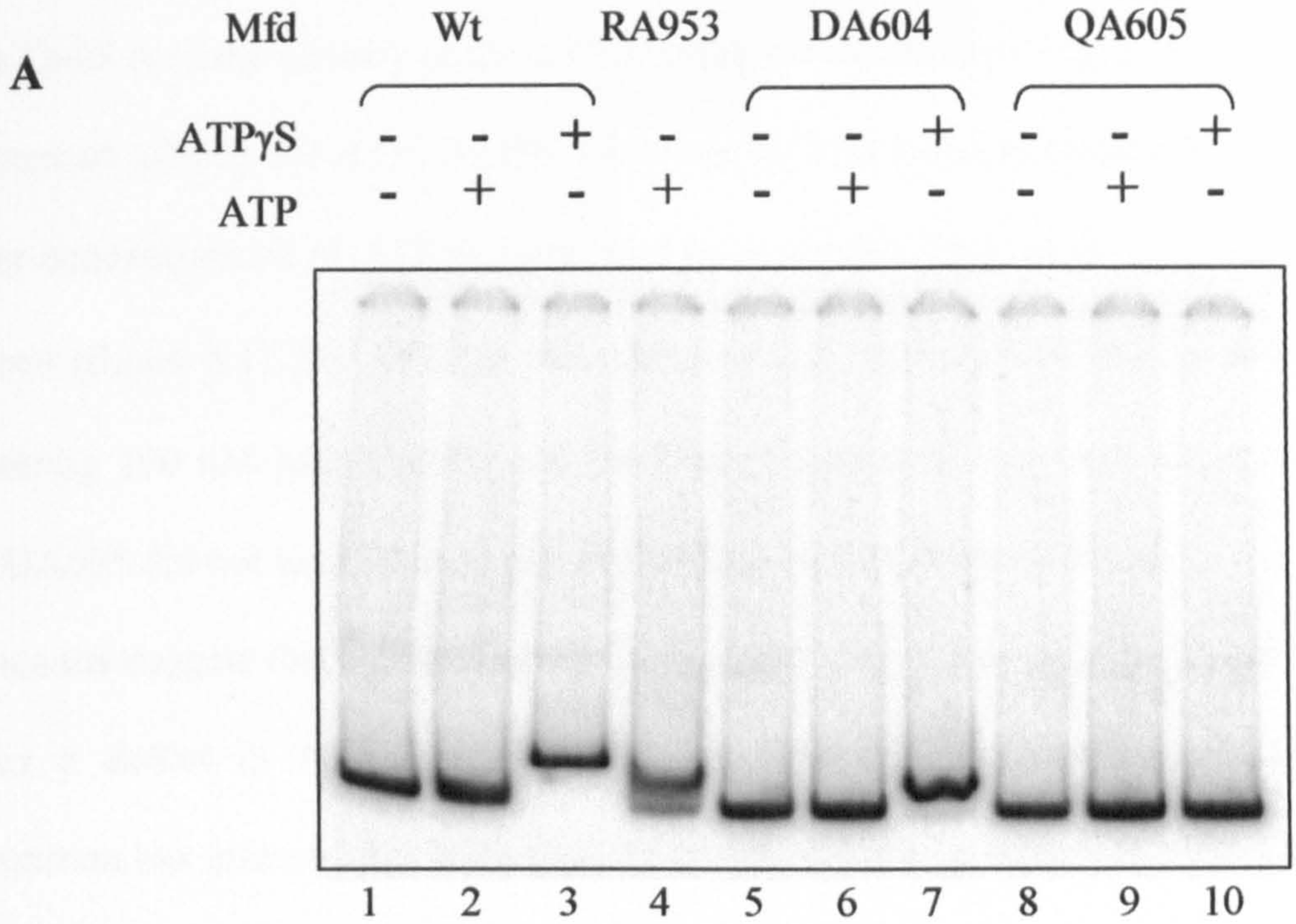


Figure 5.12. DNA binding by the Mfd proteins with substitutions in the Q-tip.

0.4 nM *Eco*RI-*Bam*HI fragment from pSRc33 was incubated with 100 nM protein in the absence of nucleotide or the presence of either ATP or ATPγS. (A) Where indicated ATP and ATPγS were present at 2 mM. (B) ATPγS was included at the concentration indicated. Samples were analysed using 8mM magnesium acetate, 5% acrylamide/TAE gels.

Since Q605 is in the vicinity of the ATP binding pocket its substitution may result in a decreased affinity for ATP. As DNA binding by Mfd requires nucleotide binding, higher concentrations of ATPγS were used to see if the DNA binding activity was restored (figure 5.12 B). ATPγS was added at 2, 4, 6, 8 and 10 mM to reactions containing 100 nM Mfd QA605 and the DNA fragment (lanes 6-10 respectively). Mfd QA605 did not bind DNA at any of these nucleotide concentrations.

The results suggest that substitution of Q605 and therefore loss of a functional Q-tip causes a defect in ATP binding in Mfd. However, the possibility that this substitution has affected the folding of the protein cannot be ruled out. Mfd DA604 was able to bind DNA in the presence of ATPγS, generating a complex that migrated with a higher electrophoretic mobility than the complex formed by the wild-type protein and resembling the complex formed by Mfd RA953 in the presence of ATP. These DNA binding properties are the same as those observed for Mfd RA929 and may indicate that both of these proteins have a lower affinity for DNA, or simply that the specific activity of the protein preparations of these two proteins is lower than the wild-type purified protein. Mfd DA604 was able to increase the efficiency of roadblock repression *in vivo* and only formed a complex with DNA in the presence of ATPγS suggesting that it is still able to hydrolyse ATP; however, the protein does not behave the same as wild type in ATP hydrolysis assays.

DISCUSSION

The key findings of studies of the DNA and nucleotide hydrolysis properties of Mfd are (i) that it is able to bind to a short (32 bp) DNA fragment but does so much more weakly than to a longer (250 bp) DNA fragment and (ii) that whilst the wild-type protein is able to utilise ATP (and to a lesser extent CTP) to displace stalled elongation complexes, this activity is abolished by substitution of the glutamine residue of the Q-tip, which appears to be crucial for nucleotide binding.

The DNA binding properties of the Mfd protein were examined *in vitro* by EMSAs on double-stranded DNA fragments 250 bp and 32 bp in length and by DNaseI footprinting. The wild-type protein forms stable complexes with DNA in the presence of ATPγS but not in the presence of ATP or in the absence of nucleotide. Therefore ATP-binding is required in order for Mfd to bind DNA, and ATP hydrolysis is associated with release of the Mfd protein from DNA.

Previous studies of the DNA binding properties of the Mfd protein had utilised DNA substrates with a minimum of 90 bp of double-stranded DNA and had suggested that the DNA may become wrapped upon binding. However, in order to mediate displacement of elongation complexes only approximately 26 bp of accessible DNA is required immediately upstream of the transcription complex. Binding of Mfd protein to a 32 bp double-stranded oligomer was more than an order of magnitude weaker than to a 250 bp DNA fragment. The affinity of Mfd for short double-stranded DNA substrates is apparently even lower than the affinity for 90 bases of single-stranded DNA (Selby and Sancar, 1995a). Binding to 250 bp of double-stranded DNA occurs with approximately the same affinity as to 95 bp of double stranded DNA. The Mfd protein from *Bacillus subtilis* displays DNA binding properties that are similar to the *E. coli* protein (Ayora *et al.*, 1996).

B. subtilis Mfd protein at 74 nM did not bind to 0.3 nM 50 bp double-stranded oligomer or four-way or Y-junction substrates in the presence of ATPγS. Binding to a 400 bp linear double-stranded DNA substrate occurred at 74 nM protein in the presence of ATPγS but not in the absence of nucleotide or in the presence of ATP. It may be that a higher concentration of *B. subtilis* Mfd protein would result in binding to a 50 bp double-stranded oligomer.

The *E. coli* Mfd protein only required a 26 bp gap between an upstream protein roadblock and the back edge of RNAP in order to displace an elongation complex (Park *et al.*, 2002). This indicates that the tighter interaction between Mfd and longer DNA fragments may not be necessary for this step in transcription-coupled repair, although the construct tested still possessed DNA downstream of the roadblock that could contribute to Mfd-DNA interactions. It is possible that affinity for the short region of DNA upstream of the elongation complex is raised by an increase in the local concentration of Mfd due to the stalled elongation complex contacting the Mfd protein and bringing it to the DNA. Tighter binding to longer sections of DNA may be required for steps in the transcription-coupled repair pathway that follow RNAP displacement e.g. if the Mfd protein remains bound to the DNA following elongation complex dissociation in order to recruit UvrA to the site of damage. This would be consistent with a requirement for 90 bp of DNA downstream of a lesion on a linear template for transcription-coupled repair.

The binding of multiple molecules of Mfd to longer fragments of DNA and also potential wrapping of DNA was examined by DNaseI footprinting. On the *Eco*RI-*Bam*HI fragment of pSRc33 two predominant regions of protection were observed on both strands. The protected region in each case was centred near a run of four Gs (or 4 Cs). Therefore, along with general protection of the fragments against DNaseI

digestion due to non-sequence specific binding of the Mfd protein, there appears to be at least a modest preference for sequences containing a run of four Gs. The region of protection that is closest to the *Bam*HI terminus actually includes two runs of 4 Gs. Protection of this region appears stronger than the region located nearer the *Eco*RI terminus (figure 5.2 A and figure 5.2 B lanes 2, 4 and 6) and this may be due to Mfd binding to either of the two G sequences, resulting in increased occupancy. Mfd protected approximately 25 bp of DNA, consistent with the 26 bp required upstream of an elongation complex for its displacement. The presence of two preferential Mfd binding positions within the fragment may suggest that the Mfd-DNA complex with lower electrophoretic mobility that was observed in EMSAs possesses Mfd bound at both of these sites and the complex with higher electrophoretic mobility has only one of these sites occupied (possibly the site nearer the *Bam*HI terminus that appears to show tighter binding). It is worth noting that the 32-mer used to examine Mfd binding to short DNA fragments did not possess a 4 G/C sequence. In addition, not all of the regions of protection observed by Selby and Sancar in their DNaseI footprinting experiments are associated with this G/C sequence, showing that it is not an absolute requirement for stable complex formation. The apparent slight sequence preference of Mfd protein could be probed further using different DNA fragments for DNaseI footprinting. There is no obvious advantage to having an Mfd binding site preference *in vivo* and any weak site preference is likely to be overcome by recruitment of Mfd to sites of stalled elongation complexes. It could be that sequences that show slightly higher affinity for Mfd are associated with a class of pause sequence so that Mfd is delivered to sites close to where RNAP becomes stalled more frequently.

Wrapping of DNA by Mfd had been proposed based on the presence of hypersensitive sites and regions of protection in previous DNaseI footprints (Selby and Sancar, 1995a). Under the conditions used for footprinting in figure 5.2 hypersensitive sites were apparent on the *Bam*HI end-labelled strand. Both this and the fact that Mfd bound a 250 bp fragment of DNA with higher affinity than a 32 bp fragment, are consistent with wrapping or bending of DNA by Mfd. If one makes the assumption that the difference in binding affinity of Mfd for short and longer DNA fragments is due to wrapping it would seem that the length of DNA required for Mfd wrapping is between 32 and 95 bp. Weaker binding to a short (35 bp) substrate in comparison to a longer substrate (147 bp) has also been observed for DNA gyrase, another protein that bends DNA (Reece and Maxwell, 1991). It could be that bending of DNA by Mfd following RNAP displacement promotes the binding of a $UvrA_2B$ complex to the nearby site of damage.

The Q-tip motif, a conserved glutamine located at the end of an α-helix upstream of helicase motif I, has been shown to be required for helicase activity of DEAD-box RNA helicases. This feature of a more extended motif in the DEAD-box RNA helicase family is present in the Mfd protein. Substitution of this glutamine residue impaired the ability of Mfd to enhance the efficiency of roadblock repression *in vivo*, presumably by reducing the ability of the protein to displace stalled elongation complexes. Substitution of a neighbouring aspartate residue that is conserved only in Mfd sequences did not impair the activity of Mfd *in vivo*. Purified Mfd DA604 protein was still able to bind to DNA in the presence of ATPγS, forming a complex equivalent to the higher mobility complex observed with wild-type Mfd, but the Mfd QA605 protein failed to bind DNA under any of the conditions that were tested, including in the presence of 10 mM ATPγS.

When the ATPase activity of the substituted proteins was examined neither protein hydrolysed ATP at a rate above the background level of hydrolysis. This was slightly unexpected for the Mfd DA604 protein since it could bind DNA in the presence of ATPγS, although not in the presence of ATP, implying that the ATP was hydrolysed to cause release of the protein from the DNA. In addition the protein was able to displace elongation complexes *in vivo* and *in vitro*, a process that requires ATP hydrolysis. As such it may be that this protein has a decreased level of ATPase activity in comparison to the wild-type that is too slow to be detected by the ATPase assay used. It may be that the Mfd QA605 protein also has residual ATPase activity since it is only partially defective in elongation complex displacement *in vivo*.

There may be other factors present *in vivo* that stimulate the ATPase activity of Mfd and therefore lessen the *in vivo* defects of Mfd proteins that are ATPase defective *in vitro*. ATPase activity is not stimulated *in vitro* by the presence of either short double-stranded DNA oligomers or *Hin*dIII-cut λ DNA or by the addition of RNAP holoenzyme (data not shown and (Selby and Sancar, 1993a). The effect of Uvr proteins on elongation complexes has not yet been tested. The residual ATP hydrolysis activity of Mfd DA604 was sufficient for elongation complex displacement in EMSAs in the presence of 2 mM ATP. This may simply reflect the fact that the EMSAs analyse a single turnover of the Mfd protein. The number of ATP molecules that need to be hydrolysed by Mfd in order to displace a single elongation complex is not known and the process could only require a modest amount of ATP hydrolysis e.g. translocation of only a few bp of DNA. If this is the case, the residual ATPase activity of Mfd DA604 may be sufficient to displace a significant proportion of the stalled complexes within 15 minutes. Elongation complex displacement by Mfd DA604 was only analysed at a single time point and it

may be that Mfd DA604 dissociates RNAP more slowly than the wild-type protein. A time course experiment could be performed to test the rate of Mfd DA604-mediated displacement.

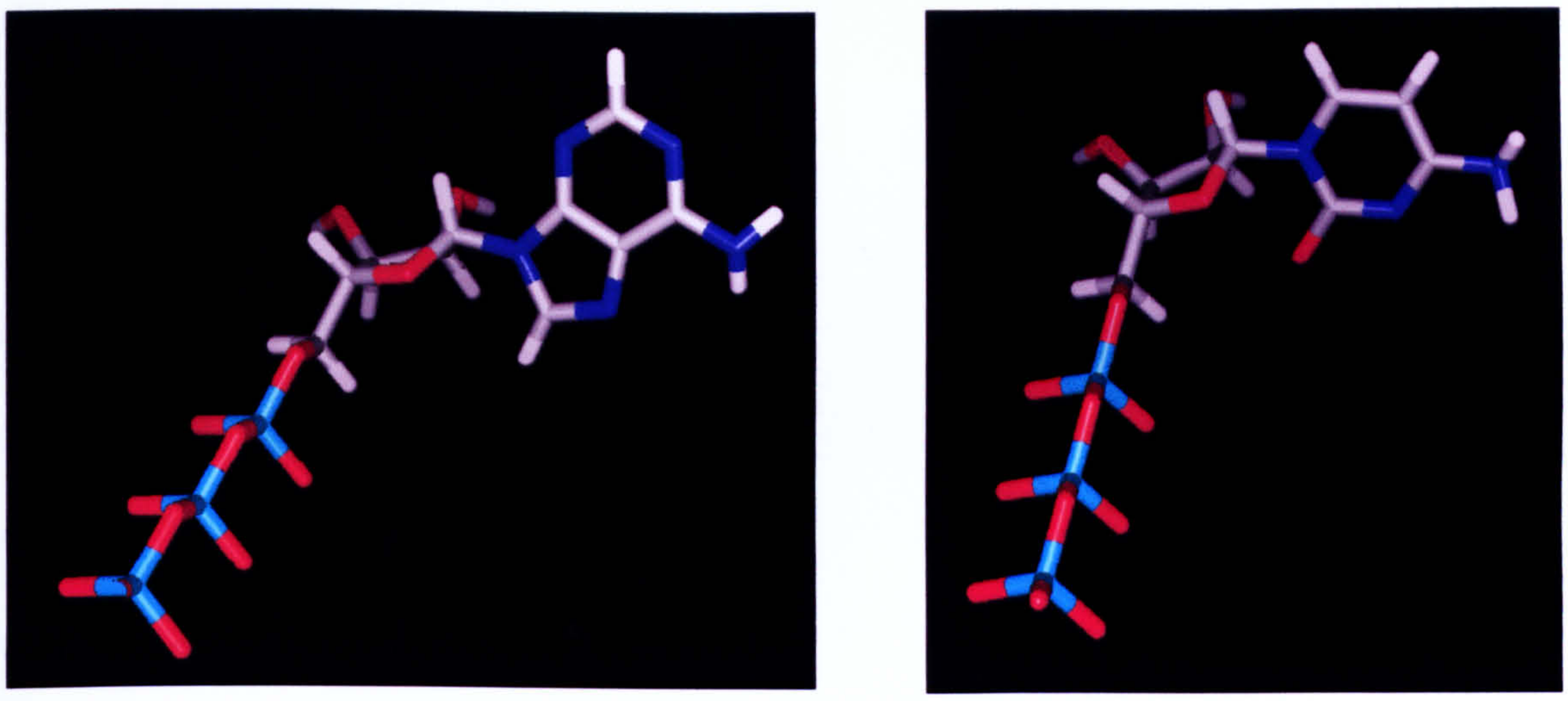
It was found that the addition of a high concentration of either CTP or UTP to an elongation complex stalled by nucleotide starvation resulted in its destabilisation in the absence of Mfd. The fact that the next nucleotide to be incorporated at the stall site was CTP may be relevant to this observation. It has been proposed that the *E. coli* elongation complex possesses a second nucleotide binding site in addition to the active site (Foster *et al.*, 2001). It is suggested that this site acts to allosterically regulate the conformational state of the transcription elongation complex and as a result regulates its activity. The model proposes that, at low nucleotide concentrations, the elongation complex can fall into an unactivated state with a slow incorporation rate. Binding of the next nucleotide to be incorporated to the allosteric site can convert the elongation complex to an activated state to permit rapid RNA synthesis. Only the nucleotide that is encoded next has this stimulatory effect. Therefore in theory, since the transcription complexes were taken to the stall site under relatively low nucleotide concentrations (2 μM UTP and 10 μM GTP), they could be in an unactivated state. Addition of a high concentration of the next nucleotide, CTP, could alter the conformation of the transcription complex into an activated state that may be less stable than the unactivated state. It is possible that the UTP nucleotide contained a small amount of CTP contamination to cause this effect upon addition of UTP.

Accounting for the decreased amount of elongation complex present following addition of CTP or UTP, the Mfd protein can use either ATP (or dATP) or less efficiently CTP (or dCTP) to displace elongation complexes. The same nucleotide

specificity was observed for the Mfd DA604 protein. Mfd QA605 was not able to utilise any of the nucleotides to displace an elongation complex *in vitro*. Due to the DNA binding, elongation complex displacement and ATP hydrolysis defects of this protein it is possible that substitution of Q605 has caused the protein to misfold or decreased its stability: however, the protein was partially able to increase roadblock repression *in vivo* in the same strain from which the protein was purified. Each of these functions of the Mfd protein requires that it is still able to bind ATP. It seems plausible that removal of this residue would result in a loss or decrease in ATP binding, since the equivalent glutamine is positioned within hydrogen bonding distance of the nucleotide base in the crystal structure of RecG. RecG is 38% identical to the Mfd protein within the helicase domains and therefore is likely to possess a similar fold.

The Mfd protein was able to utilise both ATP and CTP to cause dissociation of stalled elongation complexes (figure 5.13 A). Both of these bases possess an amine group that could hydrogen bond to the glutamine side chain (figure 5.13 B). Therefore removal of this potential contact would reduce the affinity of the protein for ATP. In addition to the conserved glutamine, a nearby aromatic residue is also highly conserved (six amino acids upstream in Mfd) and is in a position to form base stacking interactions with the nucleotide. Given that Mfd forms a stable complex with DNA in the presence of ATPγS but not in the absence of nucleotide or in the presence of ATP and also that Mfd can use either ATP or CTP to displace elongation complexes it would be of interest to examine whether Mfd can bind DNA in the presence of a non-hydrolysable CTP analogue.

A



B

Figure 5.13. Nucleotide utilisation and the Q-tip.

(A) The structures of ATP and CTP. Oxygen atoms are shown in red, nitrogen in blue and phosphorus in cyan. Both bases possess an amino group capable of forming hydrogen bonds to a glutamine side chain. (B) The Q-tip from the structure of RecG is shown indicating the potential for hydrogen bonding from the Q-tip to the adenine base and the presence of an aromatic side chain capable of forming base stacking interactions.

The Q-tip appears not only to be essential for the function of the DEAD-box RNA helicase family but is also required for Mfd-mediated elongation complex displacement. Mfd is a superfamily 2 helicase and one of the mutations found in Bloom's syndrome patients is within the Q-tip of Bloom's helicase, a member of the superfamily 2 RecQ subfamily (Bahr *et al.*, 1998). One might also expect mutation of the Q-tip of other helicases such as UvrB and RecG to affect their helicase activity. While the Q-tip has been shown to be necessary for DEAD-box helicase function this is the first report of it being required in a different helicase subfamily.

CHAPTER 6

THE RNAP INTERACTION DOMAIN OF Mfd

INTRODUCTION

Displacement of stalled elongation complexes by Mfd requires that the protein is able to bind to RNA polymerase. Interaction between a stalled elongation complex and Mfd would recruit the transcription-repair coupling factor to the halted RNAP and would permit its dissociation from the template. In binding studies performed *in vitro*, RNAP failed to bind to an Mfd affinity column but MBP-Mfd attached to amylose beads was able to pull-down RNA polymerase holo- and core enzyme. Truncations of Mfd encoding amino acids 1-571 and 379-1148 were also able to pull-down RNAP, suggesting that the RNAP contact site of Mfd falls between residues 379 and 571 (Selby and Sancar, 1995a).

A protein-protein interaction map of *H. pylori* identified an interaction between amino acids 356-488 of the *H. pylori* Mfd protein (equivalent to residues 472-603 in *E. coli*) and the β-β′ subunit of RNA polymerase, which in *H. pylori* are fused as a single polypeptide (Rain *et al.*, 2001). Residues 472-603 of *E. coli* Mfd were termed the RID (**R**NAP-**I**nteraction **D**omain), and were used as bait in a yeast two-hybrid assay against the different subunits of core RNA polymerase (Park *et al.*, 2002). The α and β′ subunits failed to interact with this region of Mfd but the β subunit bound. The region of β that interacted with the RID of Mfd was localised initially to the N-terminal 500 amino acids and then further to residues 1-142. Although the route of DNA through the elongation complex is not known for certain, parts of this region of the β subunit are close to where the upstream DNA is expected to exit the transcription complex – a suitable interaction site for a protein that requires upstream DNA for its action.

It is also worth noting that the Mfd protein from *B. subtilis* was able to displace *E. coli* RNAP stalled at a psoralen monoadduct (Ayora *et al.*, 1996), suggesting that residues critical in contacting RNAP are likely to be conserved between these two species, at least in terms of the nature of their side chain.

Determination of the residues of Mfd that contact RNAP will aid understanding of the mechanism of displacement of stalled elongation complexes by Mfd. In order to identify amino acids in the RNAP interaction domain of Mfd that are involved in the interface with RNAP and are therefore required for its displacement, the RID of Mfd has been mutated by random and site-directed mutagenesis.

RESULTS

Random mutagenesis of the Mfd RNAP interaction domain

Substitutions within the Mfd protein that abolish the interaction with RNAP would be expected to be defective in displacement of stalled elongation complexes. RNAP interaction mutants should therefore be defective in the *in vivo* roadblock repression assay described previously and an unbiased mutagenesis strategy to screen for chloramphenicol-resistant clones can therefore be used to identify residues that form part of the interface with RNAP.

Screening strategy 1

The first screening strategy that was employed (figure 6.1) was essentially the same as that used to screen for mutants of the TRG region in chapter 4. The region of *mfd* encoding amino acids 361-601 was mutated by error-prone PCR in a protocol using

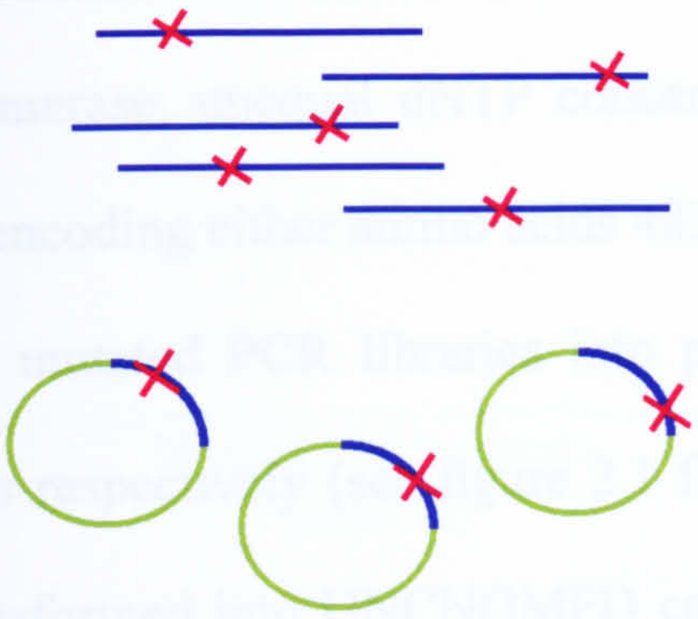


Figure 6.1. Random mutagenesis of RNAP interaction domain - screening strategy 1.

The screening strategy for identification of substitutions of Mfd that result in an inability to increase roadblock repression efficiency *in vivo* was divided into 4 stages. Stage 1 was the identification of chloramphenicol-resistant clones, stage 2 was the confirmation of this phenotype by restreaking, stage 3 was the isolation of *mfd* plasmid of the correct size and stage 4 was confirmation that the chloramphenicol-resistant phenotype was dependent on the *mfd* plasmid and DNA sequencing.

Taq polymerase, unequal dNTP concentrations and manganese (Greene J, 2002). Regions encoding either amino acids 483-582 (a) or 379-582 (b) of Mfd were cloned from the mutated PCR libraries into pETMfd2AgeI as *Age*I-*Pme*I or *Xho*I-*Pme*I fragments respectively (see figure 2.1 for restriction map). The resulting plasmids were transformed into UNCNOMFD cells containing the reporter construct pRCB-CAT and were spread onto M9 media containing kanamycin, tetracycline, ampicillin and 5 μg/ml chloramphenicol in order to identify chloramphenicol-resistant clones (figure 6.1, stage 1). The colonies that grew on chloramphenicol were restreaked to confirm their chloramphenicol-resistant phenotype and to ensure that single colonies were selected for plasmid DNA extraction (figure 6.1, stage 2). The plasmid DNA that was extracted contained two plasmids - the mutant *mfd* plasmid and the reporter plasmid. To assess the overall integrity of the mutated *mfd* plasmid an *Eco*RV digest was performed. *Eco*RV should cut once within the region of *mfd* that was subject to error-prone PCR and once within the *lacI*^{q} gene. The resulting fragments were analysed by agarose gel electrophoresis to ensure that the *mfd* plasmids that were carried further through the screen did not contain any large-scale deletions or insertions (figure 6.1, stage 3). Plasmid preparations that gave the expected restriction pattern were cut with *Nco*I, which digests only the reporter plasmid, before being transformed into UNCNOMFD cells containing pRCB-CAT, to demonstrate the dependence of the chloramphenicol phenotype on the *mfd* plasmid. The *NcoI* digested plasmid preparation was also transformed into XL1-Blue cells for extraction of pure *mfd* plasmid DNA, which could be sequenced (figure 6.1, stage 4). The number of clones that were carried through each stage of the screening was recorded (table 6.1). Initially the *Age*I-*Pme*I fragment of *mfd,* encoding amino acids

Screening stage		Fragment screened		
		*Age*I-*Pme*I Fragment (a)	*Xho*I-*Pme*I Fragment (b)	Total
	Number of colonies screened	315	3850	4165
1	Number of chloramphenicol resistant colonies	25	2270	2295
2	Number of chloramphenicol-resistant clones following restreak	16	36 out of 48	52 out of 73
3	Number of plasmids of correct size	7	1	8
4	Number of sequences determined	7	1	8
	Substitutions obtained	LR499 x2 HL527	Multiple substitutions	2 single amino acid substitutions

Table 6.1. RNAP interaction domain screening strategy 1.

The table shows the number of clones that were carried through each stage of the screening procedure shown in figure 6.1. (a) For screens of the *Age*I-*Pme*I fragment of pETMfd2AgeI 10 cycles of error-prone PCR was performed on the region encoding amino acids 361-601. The *Age*I-*Pme*I fragment encoding amino acids 483-582 was cloned into a pETMfd2AgeI vector. (b) For screening of the *Xho*I-*Pme*I fragment 8 cycles of error-prone PCR were carried out on amino acids 361-601. The *Xho*I-*Pme*I fragment encoding amino acids 379-582 was cloned into pETMfd2AgeI.

482-582, was screened using 10 cycles of error-prone PCR to generate a library of mutants. The ligation efficiency of this cloning procedure was fairly poor and as a result only 315 clones were screened by this method. Of these 25 appeared to be chloramphenicol-resistant (stage 1) and 16 were chloramphenicol-resistant upon restreaking on 5 μg/ml chloramphenicol (stage 2). Only 7 of these plasmids appeared the correct size (stage 3) and the sequences of these were determined. Two clones contained a single subsititution, LR499, although since both of these clones were isolated from the same original PCR reaction they are not independently-generated mutations. Another clone possessed a single substitution, HL527 while the remainder contained a stop codon, a frameshift and small deletions (table 6.2).

To overcome the poor ligation efficiency observed in cloning of the *Age*I-*Pme*I fragment a screen using the *Xho*I-*Pme*I fragment of *mfd* encoding amino acids 379-582 was performed using 8 cycles of error-prone PCR to generate libraries of mutants. The ligation efficiency was much improved permitting the screening of 3850 clones, of which 2270 were chloramphenicol-resistant (figure 6.1, stage 1), an unexpectedly high frequency of chloramphenicol-resistance. Forty-eight of the chloramphenicol-resistant colonies were restreaked onto plates containing 5 μg/ml chloramphenicol and 36 colonies grew (figure 6.1, stage 2). The plasmid DNA was extracted from these 36 clones but restriction mapping showed that only one of these retained the correct plasmid structure (figure 6.1, stage 3). The other plasmids all appeared to have lost large sections of DNA. Since the screen identifies loss-of-function mutants any plasmid which has rearranged and lost a large section of either the *mfd* gene or the *lacI*^{q} gene will result in the same phenotype as a single substitution that affects the interaction between Mfd and RNAP. This accounts for

A

Mutant	Base changes
Single amino acid subsitutions	**LR499** CTG→CGG
	LR499 CTG→CGG
	HL527 CAT→CTT
7 amino acid substitutions	**IF412** ATT→TTT **TI497** ACC→ATC **NS515** AAC→AGC **PS522**CCG→TCG **LP526** CTG→CCG **NI539** AAC→ATC **LF545** CTT→TTT **A418** GCC→GCT (s)
Stop codons	**Y508** TAA→TAA
	Y493 TAT→TAT **EG538** GAA→GGA
	L569 TTG→TAG **VE482** GTG→GAG **ED507** GAG→GAT **GA546** GGC→GCC
Frameshifts (Single base deletions)	**V484** GTC→GT- (fs)
	K584 AAA→AA- (fs)
	W550 TGG→TG- (fs)
	T497ACC→A-C(fs) **ML510** ATG→TTG **R492** CGT→CGC(s) **A553** GCG→GCT(s)
Deletions	Deletion of ~160 amino acids after 482 **HN527** CAT→AAT **RC552** CGC→TGC **L568** TTG→CTG (s)
	Deletion of amino acids 482-677 (fs)
	Deletion of amino acids 584-622 (fs)
Insertions	Duplication of amino acids 584-608
	Duplication of amino acids 584-608 **NT539** AAC→ACC
	Duplication of amino acids 584-608 **KR406** AAA→AGA **TS497** ACC→TCC **SI530** AGC→ATC **FS583** TTT→TCT

B

Mutant	Single amino acid subsitutions	7 subsitutions	Stop codons	Frame-shifts	Deletions	Insertions	Poor sequence
No. of clones	3	1	3	4	3	3	1

C

Base change	Frameshift (single bp deletion)	A→T	A→C	A→G	T→A	T→C	T→G	C→A	C→T	G→C	G→T
No. of changes	4	6	1	3	3	4	2	1	5	1	3
% of changes	12.1	18.2	3	9	9	12.1	6	3	15.2	3	9

Table 6.2. Mutations obtained from screening of the RNAP interaction domain of Mfd.

(A) The frameshifts, substitutions, deletions and insertions that were obtained from random mutagenesis screening of the RNAP contact domain of Mfd by screening strategies 1 and 2. Single base deletions causing frameshifts are shown as (fs) while silent base changes are shown as (s). Each line represents an individual clone. (B) Summary of the clones isolated by screening strategies 1 and 2. (C) The comparative frequency of each type of base change detected by screening strategies 1 and 2. No C to G or G to A base changes were detected.

the high frequency of chloramphenciol resistance that was observed. The sole clone that appeared to be of correct size contained multiple substitutions (table 6.2).

Screening strategy 2

When the RID of Mfd was screened using strategy 1, many clones were isolated which contained large deletions in the *mfd* plasmid. These deletions may have disrupted the $lacI^q$ gene, abolishing the formation of a roadblock upstream of the *cat* gene, resulting in a chloramphenicol-resistant phenotype. To try and reduce the high number of rearranged plasmids that were identified upon extraction of plasmid DNA, a second screening strategy was adopted (figure 6.2). The first three steps of screening strategy 2 were essentially the same as those in screening strategy 1; the region of *mfd* encoding amino acids 361-601 was subject to 5 cycles of error-prone PCR before the region encoding residues 379-582 was cloned into pETMfd2AgeI as an *Xho*I-*Pme*I fragment. The resulting plasmids were transformed into UNCNOMFD cells containing the reporter construct pRCB-CAT and the number of colonies that were chloramphenicol-resistant was recorded (figure 6.2, stage A). Overnight cultures were grown from 20 individual colonies from each transformation plate and the overnight cultures were combined in pools of 20 and the plasmid DNA extracted. The resulting DNA preparation contained 20 different potential *mfd* mutant plamids and the reporter plasmid,and was digested with *Nco*I that cut only the reporter.

Screening strategy 2 contained a secondary screen to identify plasmids that in addition to conferring chloramphenicol resistance in the primary screen, also possessed an intact $lacI^q$ gene. The strain RLG4561 has had the lactose operon deleted but contains a λ prophage with a *lacPI:lacZ* fusion and as such contains a

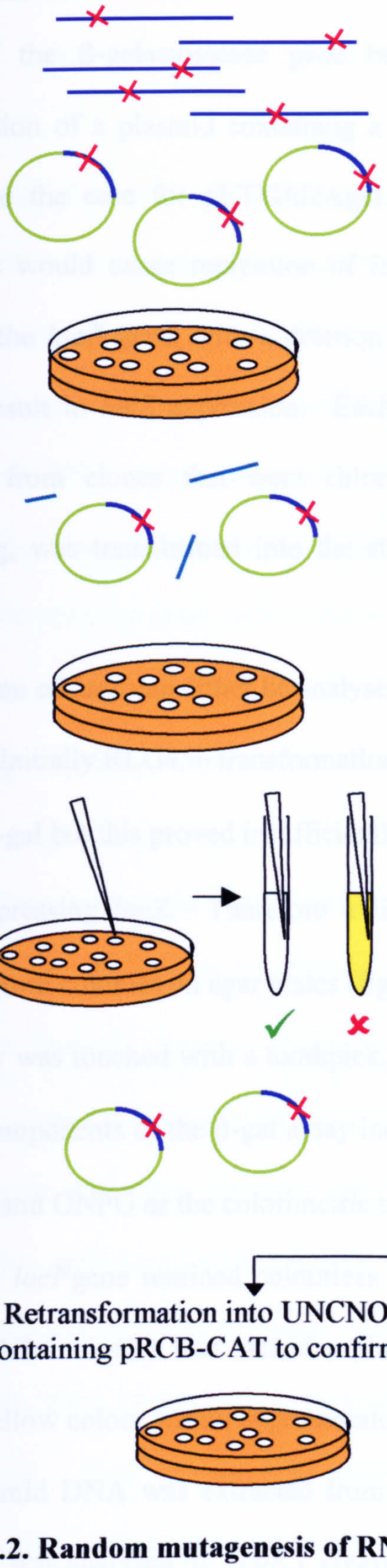


Figure 6.2. Random mutagenesis of RNAP contact region - screening strategy 2.

The screening strategy for identification of substitutions of *mfd* that result in an inability to increase roadblock repression efficiency *in vivo* was divided into 3 stages. Stage **A** was the identification of chloramphenicol-resistant clones, stage **B** was the identification of clones with intact *lacI*^{q} gene and stage **C** was the isolation of *mfd* plasmid of the correct size for sequencing.

copy of the β-galactosidase gene but no Lac repressor protein. However, introduction of a plasmid containing a fully functional copy of the *lacI*^{q} gene, as would be the case for pETMfd2AgeI plasmids that did not contain large-scale deletions would cause repression of *lacZ* expression. Introduction of plasmids without the *lacI*^{q} gene, with a deletion in *lacI*^{q} or with a non-functional *lacI*^{q}gene would result in *lacZ* expression. Each of the mixed plasmid DNA preparations, isolated from clones that were chloramphenicol-resistant in the first stage of screening, was transformed into the strain RLG4561, to carry out the secondary screen.

Expression of *lacZ* can either be analysed using X-gal plates or by liquid assay using ONPG. Initially RLG456 transformations were plated onto agar plates containing 40 μg/ml X-gal but this proved insufficiently sensitive to clearly identify which colonies were expressing *lacZ*. Therefore a liquid β-galactosidase assay was performed directly from colonies on agar plates (figure 6.2, stage B).

A colony was touched with a toothpick, which was then placed in a tube containing all the components of the β-gal assay including chloroform and SDS to permeabilise the cells and ONPG as the colorimetric substrate. Colonies with plasmids expressing an intact *lacI*^{q}gene remined colourless when assayed in the liquid β-galactosidase assay, while colonies that lacked a plasmid expressing the *lacI*^{q} gene produced a strong yellow colour within approximately two minutes at room temperature.

The plasmid DNA was extracted from those clones that possessed an intact *lacI*^{q} gene and was analysed by *Eco*RV cleavage and gel electrophoresis (figure 6.2, stage C). Those plasmids that produced the correct restriction map were transformed into UNCNOMFD cells containing the reporter pRCB-CAT and were plated on chloramphenicol plates to confirm that the phenotype was dependent on the *mfd*

plasmid. The DNA sequence of plasmids that conferred an *mfd*⁻ phenotype was determined.

The number of clones carried through each stage of the screening strategy is shown in table 6.3. In stage A approximately 27800 clones were screened of which 3077 clones were able to grow on chloramphenicol. Overnight cultures were grown of 500 of these colonies and the plasmid DNA was extracted in pools of 20 generating 25 plasmid DNA preparations that contained a mixture of mutated *mfd* plasmids. These plasmid preparations were transformed into RLG4561 cells for the secondary screen. 20 colonies from each of these transformations (500 in total) were tested in the colony β-gal assay and only 48 of those assayed (9.6%) remained colourless, indicating that the plasmid contained an intact *lacI*^{q} gene (table 6.3, stage B). The majority of the plasmids presumably contained deletions that disrupted the *lacI*^{q} gene. When the plasmid DNA was extracted from the 48 clones that were positive in the secondary screen only 10 (20.8%) gave the expected *Eco*RV restriction pattern (table 6.3, stage C), however those plasmids that were incorrect had not undergone the same dramatic deletions that were observed in the previous screen. The majority of plasmids that were incorrect only cut once rather than twice with *Eco*RV, presumably indicating that the *Eco*RV site within the *lacI*^{q} gene was intact and that small deletions or rearrangements within the *mfd* gene had occurred. The DNA sequences of the 10 plasmids that produced the expected restriction fragments were determined. Stop codons were present in two of the sequences, 3 contained frameshifts, 3 contained insertions, one contained a deletion and one did not sequence (table 6.2). The most common base changes were A to T, C to T, T to C and single base deletions causing frameshifts, similar to the distribution of base changes observed in screening of the TRG motif. Despite the failure to isolate any

Screening stage		Clones carried through each stage
	Number of colonies screened	27800
A	Number of chloramphenicol-resistant colonies	3077
B	Number of clones expressing *lacI*^{q}	48 out of 500
C	Number of plasmids of correct size	10
	Single amino acid substitutions obtained	0

Table 6.3. RNAP interaction domain screening strategy 2.

The table shows the number of clones that were carried through different stages of the screening protocol. The mutated libraries were created by error-prone PCR and cloning the *Xho*I-*Pme*I fragment of *mfd,* encoding amino acids 379-582, into a pETMfd2AgeI vector.

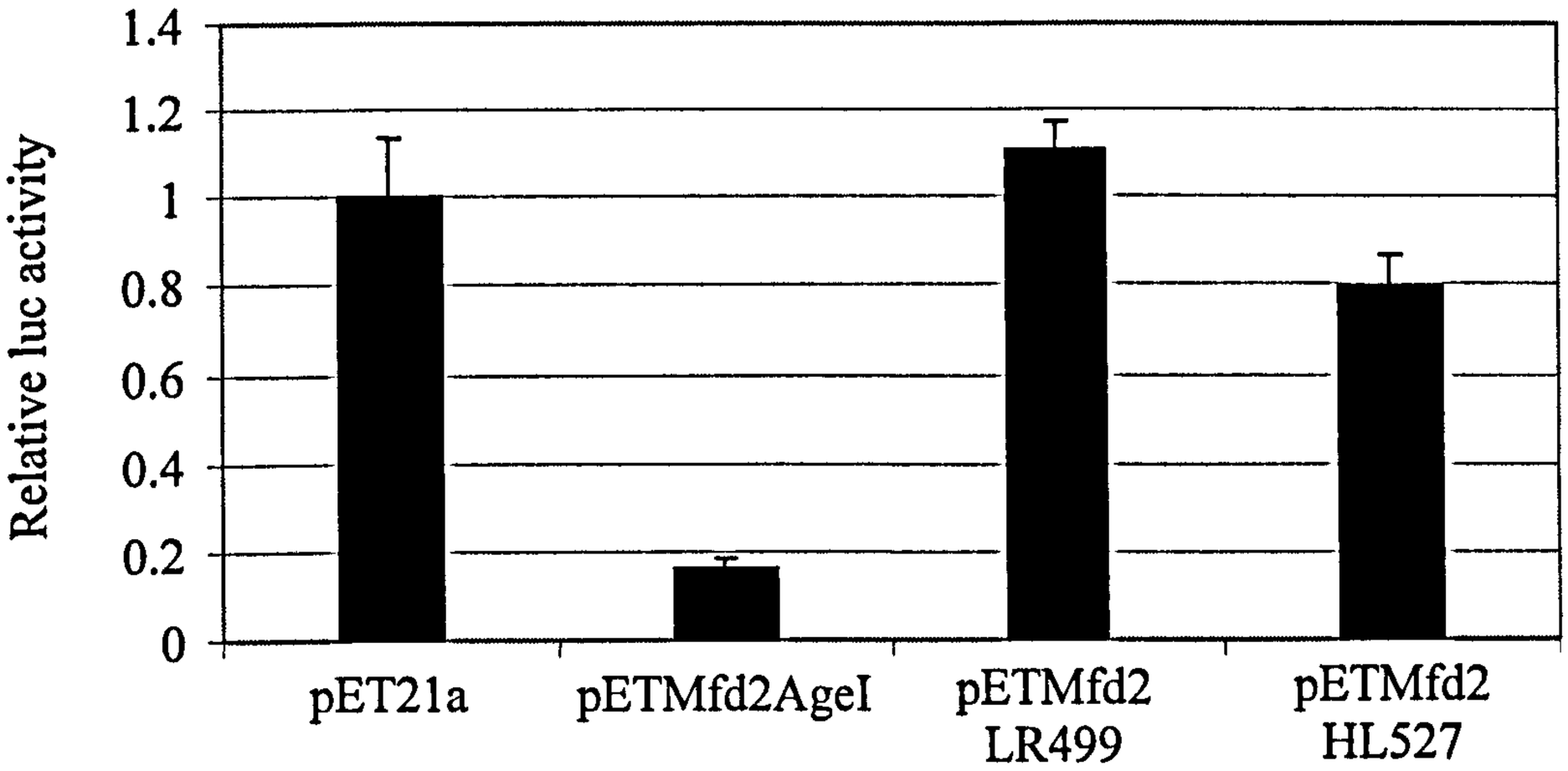
single amino acid substitutions the introduction of a secondary screen improved the screening process by permitting a greater number of clones to be examined for intact *mfd* plasmids. 500 clones were tested in the secondary screen allowing elimination of all but 48 prior to plasmid extraction. Only 73 clones were examined for intact *mfd* plasmids in screening strategy 1 since plasmid DNA extraction is a more labour intensive screening procedure than the colony β-galactosidase assay that was introduced in strategy 2. In addition, following the secondary screening procedure all clones contained a mutation within the region of *mfd* that was subjected to error-prone PCR, even though these were stop codons, deletions, frameshifts and insertions rather than single amino acid substitutions. Following plasmid DNA extraction the percentage of plasmids that possessed the correct restriction pattern was not dramatically different for the two screening procedures (15.3% using strategy 1 and 20.8% using strategy 2). This is despite the fact that a large number of incorrect plasmids had been eliminated prior to this step in strategy 2. This may be accounted for by the fact that fewer error-prone PCR cycles were performed in screen 2 and therefore the frequency of substitutions that abolished interaction with RNAP would have been lower and therefore scarcer in relation to the large number of plasmid deletions that were observed.

Overall two single substitutions were isolated from screening of the RID of *mfd* using random mutagenesis. L499 of *E. coli* Mfd is conserved in *B. subtilis* Mfd while H527 of *E. coli* Mfd is replaced by an aspartate residue in *B. subtilis*. Residues required for the interaction of Mfd with RNAP would be expected to be defective in elongation complex displacement. The ability of these mutants to displace elongation complexes *in vivo* in the luciferase roadblock repression assay was quantitated.

UNCNOMFD cells were transformed with the roadblock reporter construct pRCBKA4 and either the empty vector pET21a, pETMfd2AgeI, pETMfd2LR499 or pETMfd2HL527. The cultures were grown to mid-exponential phase and their luciferase activity was determined (figure 6.3). Cells containing pETMfd2AgeI encoding wild-type Mfd plasmid possessed 6.25 lower luciferase activity than cells containing pET21a. Cells transformed with either pETMfd2LR499 or pETMfd2HL527 had luciferase activity comparable to those containing pET21a. Therefore the efficiency of roadblock repression had not been increased in cells containing plasmids encoding Mfd LR499 or Mfd HL527. An attempt to purify HL527 protein from UNCNOMFD cells failed due to a lack of protein expression. Therefore the failure of pETMfd2HL527 to increase the efficiency of roadblock repression is due to an absence of Mfd protein, either because the substitution makes the protein less stable or perhaps the plasmid had acquired a mutation within the *mfd* promoter that affects expression. Mfd LR499 was expressed and successfully purified from UNCNOMFD cells and therefore the failure of this Mfd protein to increase the efficiency of roadblock repression *in vivo* could be due to a defect in RNAP binding or alternatively a defect in ATP hydrolysis, DNA binding or RNAP displacement.

Site-directed mutagenesis of the Mfd RNAP interaction domain

As only one substitution that affected the ability of Mfd to increase the efficiency of roadblock repression was isolated by random mutagenesis screening, site-directed mutagenesis of the RID of Mfd was also carried out. Amino acids 472-603 of *E. coli* Mfd interacted with the β subunit of RNAP in a yeast two-hybrid assay (Park *et al.*, 2002) and amino acids 1-571 bound to RNAP in pull-down assays (Selby and Sancar, 1995a) so it appears that the amino acids responsible for RNAP interaction



Mfd plasmid	Normalised luciferase activity relative to pET21a	Fold Mfd effect (Mfd plasmid/pET21a)
pETMfd2AgeI	0.16+/-0.02	6.25
pETMfd2LR499	1.11+/-0.06	0.9
pETMfd2HL527	0.79+/-0.07	1.27

Figure 6.3. The effect of LR499 and HL527 substitutions on roadblock repression.

UNCNOMFD cells containing the roadblock reporter construct pRCBKA4 were transformed with either the empty vector pET21a or plasmids expressing wild-type Mfd or Mfd proteins with LR499 or HL527 substitutions. Cells were grown at 37°C in M9 media containing kanamycin, tetracycline and ampicillin until mid-exponential phase. Cells were harvested and their luciferase activity was determined. Luciferase activity was normalised by dividing the luciferase reading by the A_{600} at harvesting. Relative luciferase values were calculated by taking the normalised luciferase activity of pET21a-containing cells as 1 and are shown with s.d. Assays were performed in triplicate.

are located in the region 472-571. Two criteria were applied to select residues in this region to be targeted by site-directed alanine mutagenesis. Firstly, the *B. subtilis* Mfd protein is able to displace *E. coli* RNAP and therefore the residues (or at least the nature of the side chains) involved in contacting *E. coli* RNAP should be conserved between the Mfd proteins from these species. Secondly, Dr. A. Smith in our lab. has isolated mutants of *rpoB* that render RNA polymerase resistant to Mfd-mediated displacement and rescue from a backtracked state (personal communication). The residues that appear to form an Mfd contact patch in the β subunit are βI117, βK118 and βE119. Substitution of E119 had the greatest effect; an alanine substitution at this position abolished elongation complex displacement by wild-type Mfd. Given that two of the critical β subunit residues are charged it is likely that charged residues in the Mfd protein are involved in contacting this patch of RNAP. In support of this 1 M salt eluted RNAP from MBP-Mfd bound to amylose beads (Selby and Sancar, 1995a).

Amino acids 472-571 of *E. coli* Mfd were aligned by Clustal W with the equivalent region of *B. subtilis* Mfd (amino acids 495-594) (figure 6.4). The Mfd proteins from these two species are 40% identical within this region. Conserved charged amino acids, including histidine, were identified and replaced by alanine residues in pETMfd2AgeI. In addition, amino acids that aligned with residues of the same charge i.e. aspartate and glutamate or lysine and arginine were substituted with alanine in pETMfd2AgeI. This led to the generation of 15 site-directed mutants encoding the following *E. coli* Mfd proteins: - RA472, EA476, HA485, HA488, RA492, EA500, EA507, KA518, RA531, KA544, RA552, RA554, KA556, DA563 and EA567.

```
                472  476          485 488  492         500       507
E.coli          RNLAELHIGQPVVHLEHGVGRYAGMTTLEAGGITGEYLMLTY
B.subtilis      KSYSELQIGDYVVHINHGIGKYLGIETLEINGIHKDYLNIHY
                :. :**:**: ***:**:*:* *: *** .**  :** : *

                     518             531             544       552 554
E.coli          ANDAKLYVPVSSLHLISRYAGGAEENAPLHKLGGDAWSRARQ
B.subtilis      QGSDKLYVPVEQIDQVQKYVGSEGKEPKLYKLGGSEWKRVKK
                .. ******.:. :.:*.*.  ::. *:****. *.*.::

                556       563   567
E.coli          KAAEKVRDVAAELLDI
B.subtilis      KVETSVQDIADDLIKL
                *.  .*:*:* :*:.:
```

Figure 6.4. Alignment of the RNAP contact region of Mfd from *E. coli* and *B. subtilis*.

Residues 472-571 of Mfd from *E. coli* were aligned with residues 495-594 of Mfd from *B. subtilis* by ClustalW. The proteins are 40% identical in this region. Conserved charged amino acids are shown in bold and numbers refer to the position in *E. coli*. Positively charged amino acids are shown in red, negatively charged residues in blue and histidine in purple. In addition the positions of the two substitutions isolated by random mutagenesis are shown in green. Stars indicate identity, two dots conservation and one dot indicates similarity.

Substitutions that disrupt the interaction between Mfd and RNAP are expected to fail to displace elongation complexes stalled at the Lac repressor roadblock and therefore will result in high *luc* or *cat* expression in the roadblock repression assay, in the case of the *cat* reporter system this will cause chloramphenicol-resistance. Each of the site-directed mutants was tested for its ability to displace stalled elongation complexes *in vivo* using both the CAT reporter system and the luciferase reporter system.

Firstly each of the site-directed mutant plasmids, in addition to pET21a and pETMfd2AgeI, was transformed into UNCNOMFD cells containing the *cat* reporter pRCB-CAT. The ability of each of the mutants to confer chloramphenicol-resistance was examined by spreading the transformations onto M9 agar plates containing kanamycin, tetracycline, ampicillin and 5 μg/ml chloramphenicol. Transformation with pET21a conferred chloramphenicol-resistance on the cells while cells containing pETMfd2AgeI were chloramphenicol sensitive. Only two of the fifteen site-directed mutants, pETMfd2HA485 and pETMfd2RA554 conferred chloramphenicol-resistance (figure 6.5).

Each of the site-directed mutants was also assayed for Mfd activity in liquid culture using the luciferase roadblock repression system. UNCNOMFD cells containing the reporter construct pRCBKA4 were transformed with pET21a, pETMfd2AgeI or each of the site-directed charge mutant plasmids. Cultures were grown to mid-exponential phase and their luciferase activity was determined (figure 6.5).

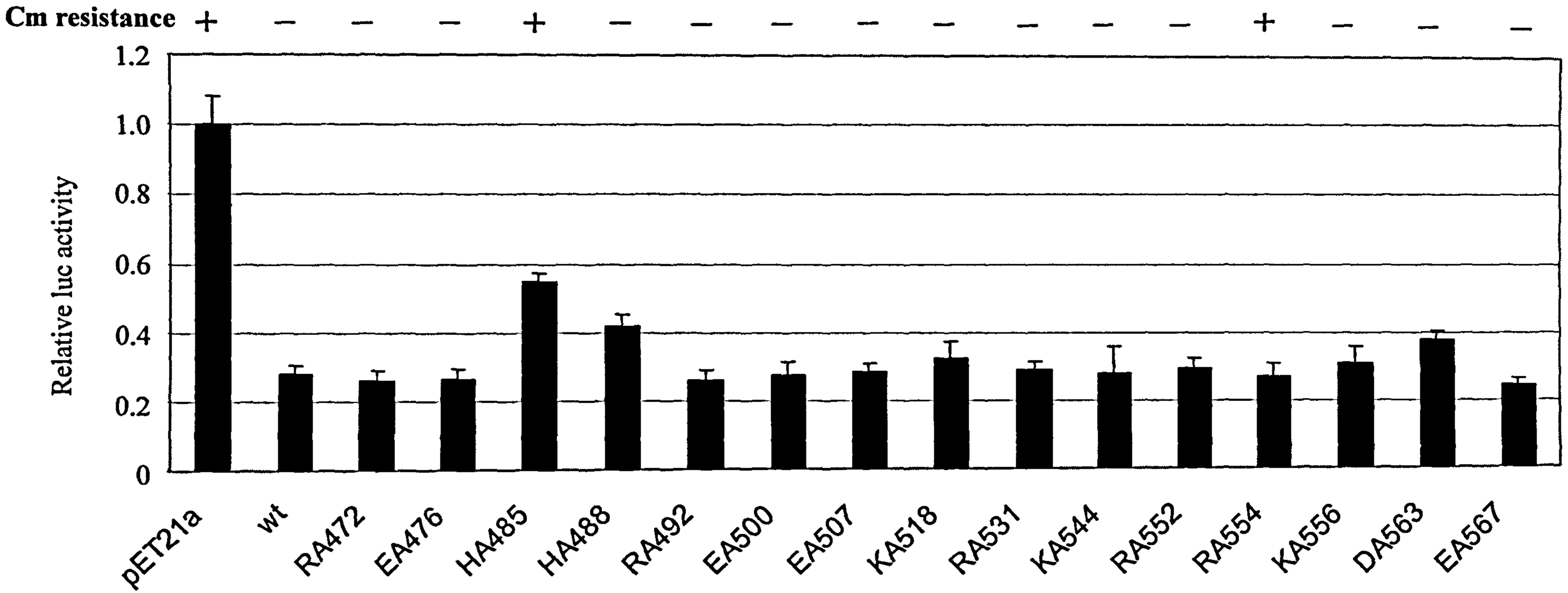
Luciferase activity was approximately 3.6-fold lower in cells containing wild-type Mfd than in cells without Mfd. The luciferase activity in cells containing wild-type Mfd was not statistically significantly different from cells containing RA472, EA476, HA488, RA492, EA500, EA507, KA518, RA531, KA544, RA552, RA554,

Figure 6.5. Effect of Mfd RID site-directed alanine substitutions on roadblock repression *in vivo*.

UNCNOMFD cells were transformed with each of the alanine substitution pETMfd2AgeI plasmids along with the empty vector pET21a and pETMfd2AgeI and either the *cat* reporter construct pRCB-CAT or the *luc* reporter construct pRCBKA4.

Cells were tested in the *cat* roadblock repression system for resistance to chloramphenicol by spreading the transformations onto M9 agar plates containing kanamycin, tetracycline, ampicillin and 5 μg/ml chloramphenicol.

Luciferase assays were performed from liquid cultures grown to mid-exponential phase. Cells were harvested and lysed and the luciferase activity determined. The luciferase activity was normalised to the optical density at which the cultures were harvested and then expressed relative to the activity of cells containing the negative control, pET21a and are shown with s.d. Assays were performed in triplicate.



Substitution	Wt	RA 472	EA 476	HA 485	HA 488	RA 492	EA 500	EA 507	KA 518	RA 531	KA 544	RA 552	RA 554	KA 556	DA 563	EA 567
Relative luc activity	0.278 +/- 0.025	0.377 +/- 0.010	0.266 +/- 0.032	0.546 +/- 0.019	0.420 +/- 0.016	0.259 +/- 0.032	0.267 +/- 0.038	0.283 +/- 0.028	0.342 +/- 0.050	0.288 +/- 0.017	0.277 +/- 0.042	0.293 +/- 0.028	0.270 +/- 0.034	0.308 +/- 0.056	0.378 +/- 0.012	0.244 +/- 0.024
Fold Mfd effect	3.6	2.7	3.8	1.8	2.4	3.9	3.7	3.5	2.9	3.5	3.6	3.4	3.7	3.2	2.6	4.0

KA556, DA563 or EA567 substituted Mfd proteins. The only substitution that resulted in luciferase activity statistically higher than the wild-type Mfd protein was HA485.

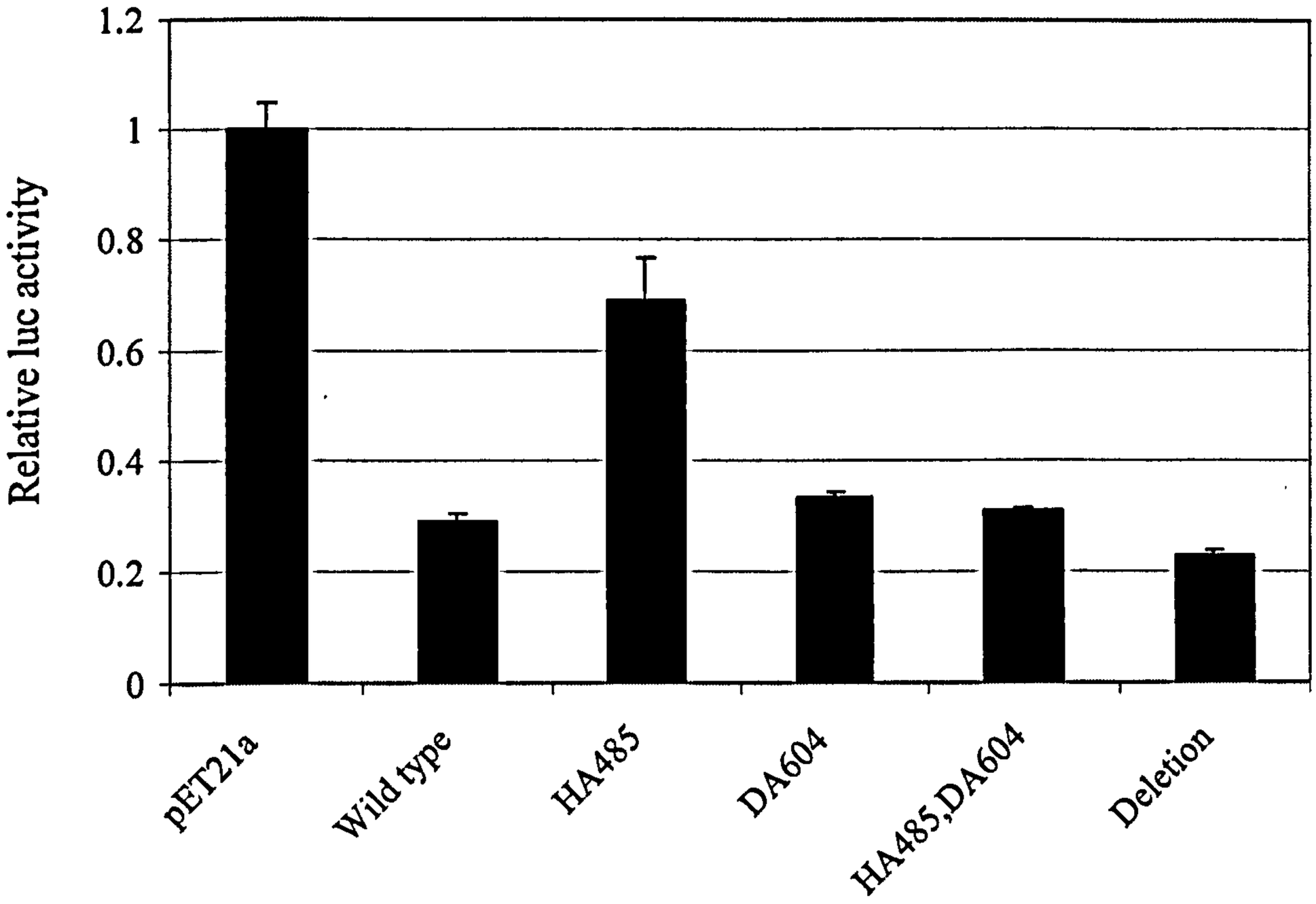
Therefore Mfd HA485, which resulted in chloramphenicol-resistance of *mfd*⁻ cells containing the reporter construct pRCB-CAT, was also defective in luciferase assays in liquid culture, whilst Mfd RA554 behaved like the wild-type protein in luciferase assays but was chloramphenicol-resistant in the *cat* system. It was shown in chapter 4 that chloramphenicol-resistance on solid media is a more sensitive assay of Mfd function that assays performed in liquid culture. The simplest explanation for the defects of Mfd HA485 and Mfd RA554 proteins *in vivo* is that they are defective in RNAP displacement due to a defect in RNAP recruitment. Alternatively the substitutions could affect the folding or stability of the protein or affect another function of Mfd such as DNA-binding, ATP hydrolysis or RNAP displacement.

During the generation of the conserved charge alanine substitutions described above two additional mutants were created serendipitously. The first of these mutants contained two amino acid substitutions: HA485, which was shown to affect RNAP displacement *in vivo,* and DA604. D604 is conserved across all Mfd protein sequences and is located next to the glutamine of the Q-tip that appears likely to be involved in ATP binding (see chapter 5). An Mfd protein containing the single substitution DA604 was able to displace stalled elongation complexes both *in vitro* and *in vivo* and bound DNA in an ATPγS-dependent manner but displayed a defect in ATP hydrolysis (see chapter 5). The second mutant contained an in-frame deletion of amino acids 362-468 along with LP361, RA472 and RA531 substitutions. It therefore does not contain the first 89 amino acids of the RNAP contact domain that was initially identified by pull-down assays (379-571) (Selby and Sancar,

1995a). RA472 and RA531 substitutions did not affect the ability of Mfd to displace stalled elongation complexes *in vivo* in either the *luc* or *cat* roadblock repression assays. The ability of the Mfd HA485 and Mfd DA604 single substitutions, in addition to the Mfd HA485, DA604 double mutant and the deletion mutant containing LP361, RA472 and RA531 substitutions, to displace RNAP in the *luc* roadblock system was examined.

UNCNOMFD (*mfd*) cells containing the roadblock reporter construct pRCBKA4 were transformed with the empty vector pET21a, pETMfd2AgeI expressing wild-type Mfd or derivatives of pETMfd2AgeI encoding Mfd DA604 and Mfd HA485, the double substitution Mfd HA485, DA604 or finally the 362-468 amino acid deletion. Cultures were grown to mid-exponential phase and the luciferase activity was measured (figure 6.6).

The normalised luciferase activity was 3.5-fold lower in cells containing wild-type Mfd protein than in cells transformed with the empty vector pET21a. Luciferase activity of cells containing Mfd DA604 was the same as that of cells containing the wild-type protein (Mfd DA604 also behaved like the wild-type protein in the *cat* roadblock repression assay in chapter 5). As observed previously, cells containing Mfd HA485 possessed an increased level of luciferase activity in comparison with cells expressing the wild-type protein, indicating that substitution of HA485 caused a defect in RNAP displacement *in vivo*. However the double substitution Mfd HA485, DA604 behaved like the wild-type Mfd protein and was able to increase the efficiency of roadblock repression *in vivo*. The Mfd protein in which amino acids 362-468 had been deleted (in addition to LP361, RA472 and RA531 substitutions) also behaved like wild-type protein in the roadblock repression assay.



Mfd substitution	Relative *luc* activity	Fold Mfd effect (-Mfd plasmid/+Mfd plasmid)
Wild-type	0.289 +/- 0.015	3.5
HA485	0.690 +/- 0.077	1.4
DA604	0.333 +/- 0.012	3
HA485, DA604	0.309 +/- 0.009	3.2
362-468 deletion	0.228 +/- 0.013	4.4

Figure 6.6. The effect of Mfd HA485, DA604 double substitution and deletion of amino acids 362-468 on Mfd activity *in vivo*.

UNCNOMFD (*mfd*) cells containing the roadblock repression reporter construct pRCBKA4 were transformed either with pET21a, pETMfd2AgeI or derivatives of pETMfd2AgeI encoding Mfd with HA485 and DA604 single substitutions or a HA485, DA604 double substitution or with deletion of amino acids 362-468 (in addition to LP361, RA472 and RA531 substitutions). Cultures were grown to mid-exponential phase and the luciferase activity was determined. Luciferase activity was normalised by dividing by the optical density at harvesting and is expressed relative to the normalised luciferase activity of cells containing pET21a. Values shown are the average of three experiments and are shown with s.d.

Therefore although substitution of H485 with alanine resulted in an Mfd protein that was defective in elongation complex displacement *in vivo,* when this substitution was combined with the substitution DA604, the RNAP dissociation activity of the Mfd protein was restored. It also appears that amino acids 362-468 are not essential for the RNAP displacement activity of Mfd *in vivo*. This suggests that this region is not involved in RNAP recruitment and that there is not a contact necessary for displacement formed between amino acids in this domain and RNAP. This region is part of the RNAP interaction domain identified by RNAP pull-down assays (amino acids 379-571) (Selby and Sancar, 1995a) but is not part of the shorter RNAP interaction domain identified by yeast two-hybrid (amino acids 472-603) (Park *et al.*, 2002), supporting the conclusion that the residues essential for RNAP contact are located in the region 472-571.

Effect of substitutions in the RID on Mfd activity *in vitro*

The LR499, HA485 and RA554 substitutions are located within the proposed RNAP interaction domain of Mfd and therefore their inability to displace stalled elongation complexes is expected to be caused by a defect in RNAP recruitment. However, their defects could also be caused by a defect in DNA binding, ATP hydrolysis, coupling ATP hydrolysis to RNAP displacement or a specific deficiency in elongation complex dissociation. To examine which of these functions of Mfd were affected by the substitutions, the properties of the purified Mfd proteins were tested *in vitro*. Although the HA485 protein was expressed (data not shown), it could not be purified by the protocol (using BlueSepharose, heparin and MonoQ columns) that had been used to purify the wild-type protein and other mutant Mfd proteins. However, Mfd LR499 and Mfd RA554 were purified successfully and these proteins

were tested for their ability to hydrolyse ATP, bind DNA in an ATPγS-dependent manner and to displace elongation complexes *in vitro*.

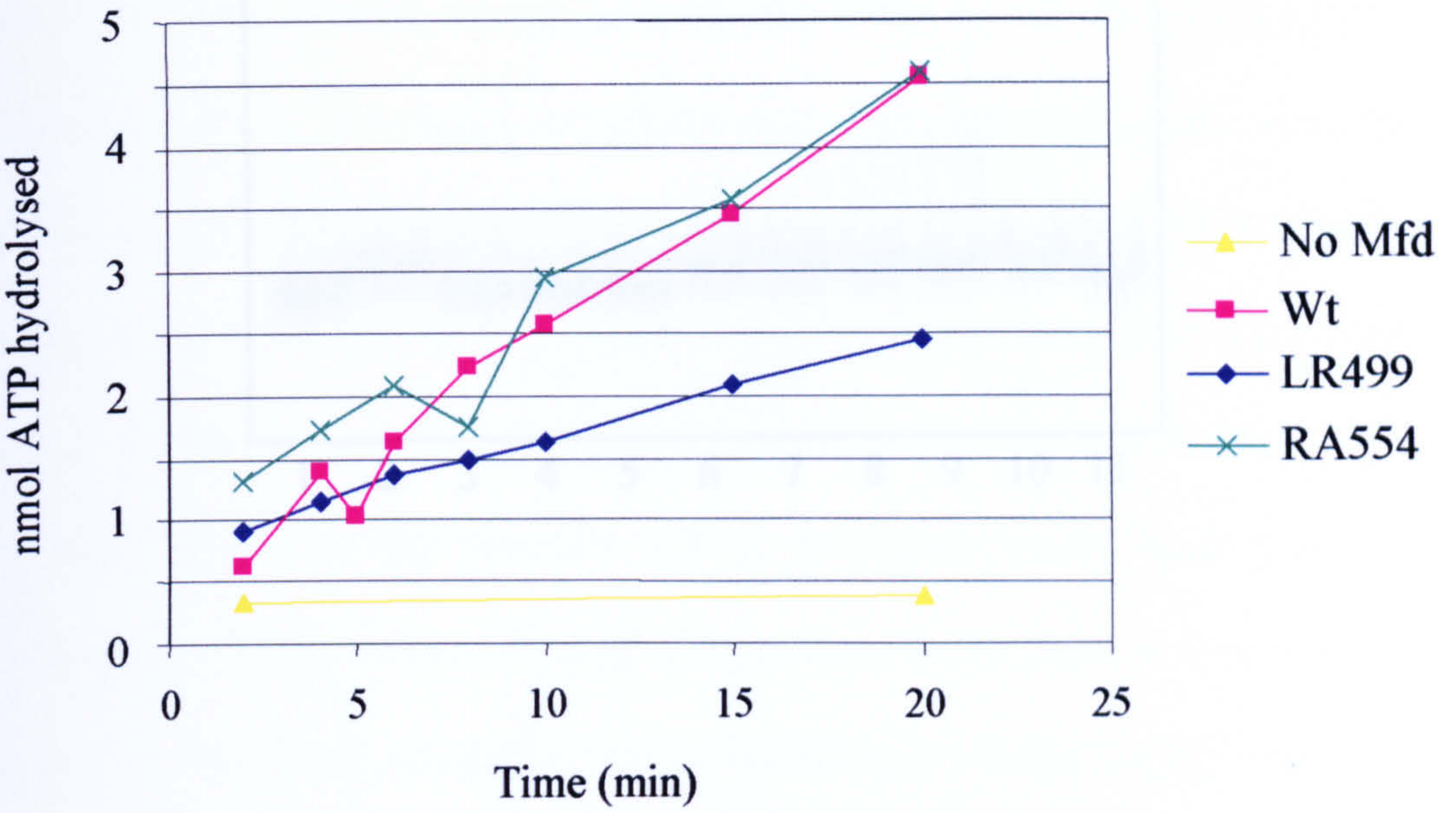
ATP hydrolysis by Mfd proteins containing substitutions in the RID

ATP (or CTP) hydrolysis is necessary for displacement of stalled elongation complexes and therefore Mfd with defective ATPase activity would be unable to displace RNAP in the *in vivo* roadblock repression assays. To establish whether the substitutions that are expected to interfere with the interaction between Mfd and RNAP caused a defect in ATP hydrolysis, k_{cat} values were determined for the purified proteins.

The hydrolysis of ATP by 1 μM Mfd protein was followed over time (figure 6.7). Mfd RA554 hydrolysed ATP with a k_{cat} of 18 min^{-1} and Mfd LR499 with a k_{cat} of 8 min^{-1}. Although the k_{cat} value is slightly lower than that of the wild-type protein (22 min^{-1}) at such low rates of ATP turnover this variation falls within the error of the experiment. Therefore substitution of L499 and R554 did not abolish ATP hydrolysis activity and this suggests the substitutions did not grossly affect the fold of this region of the protein.

DNA binding by Mfd proteins containing substitutions in the RID

DNA upstream of an elongation complex is required for its displacement and therefore the failure of Mfd LR499 and Mfd RA554 proteins to displace RNAP *in vivo* could be caused by a defect in DNA binding. In order to determine whether substitution of L499 and R554 affect the ATPγS-dependent DNA binding activity of Mfd, the purified proteins were incubated with the 250 bp *Eco*RI-*Bam*HI fragment of pSRc33 either in the absence of nucleotide or in the presence of ATP or ATPγS and the resulting complexes were analysed by EMSA (figure 6.8).

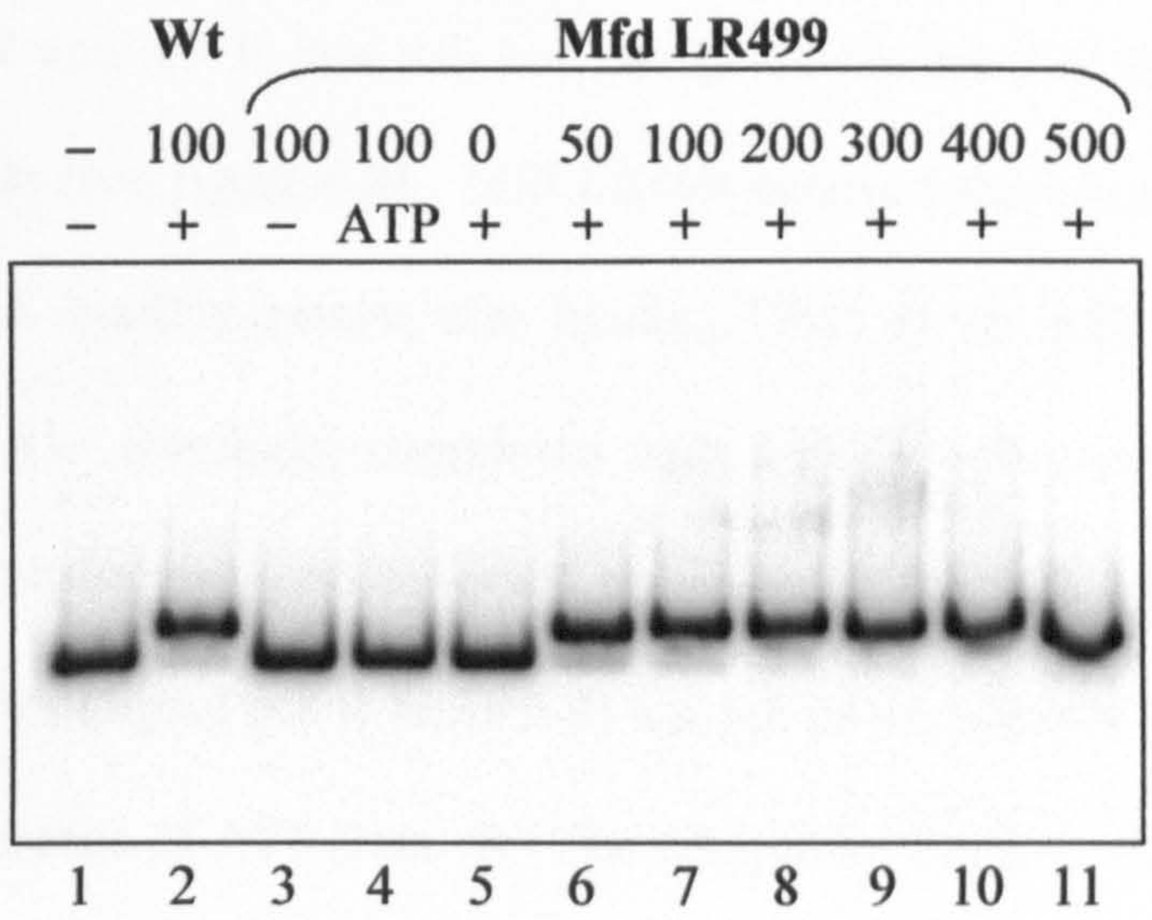


Mfd protein	k_{cat}
Wild type	22 +/- 3 min^{-1}
LR499	8.32 +/- 0.37 min^{-1}
RA554	18.3 +/- 0.64 min^{-1}

Figure 6.7. ATPase activity of LR499 and RA554 Mfd proteins.

1 μM purified Mfd protein was incubated in a 10 μl reaction with 2 mM ATP and 1 μCi [γ-^{32}P] ATP at 37°C. Unhydrolysed ATP was adsorbed using activated charcoal and Cerenkov counting was used to quantify hydrolysed ATP at different time points. The graph shown is the average of two experiments and k_{cat} values were calculated with standard deviation.

A



B

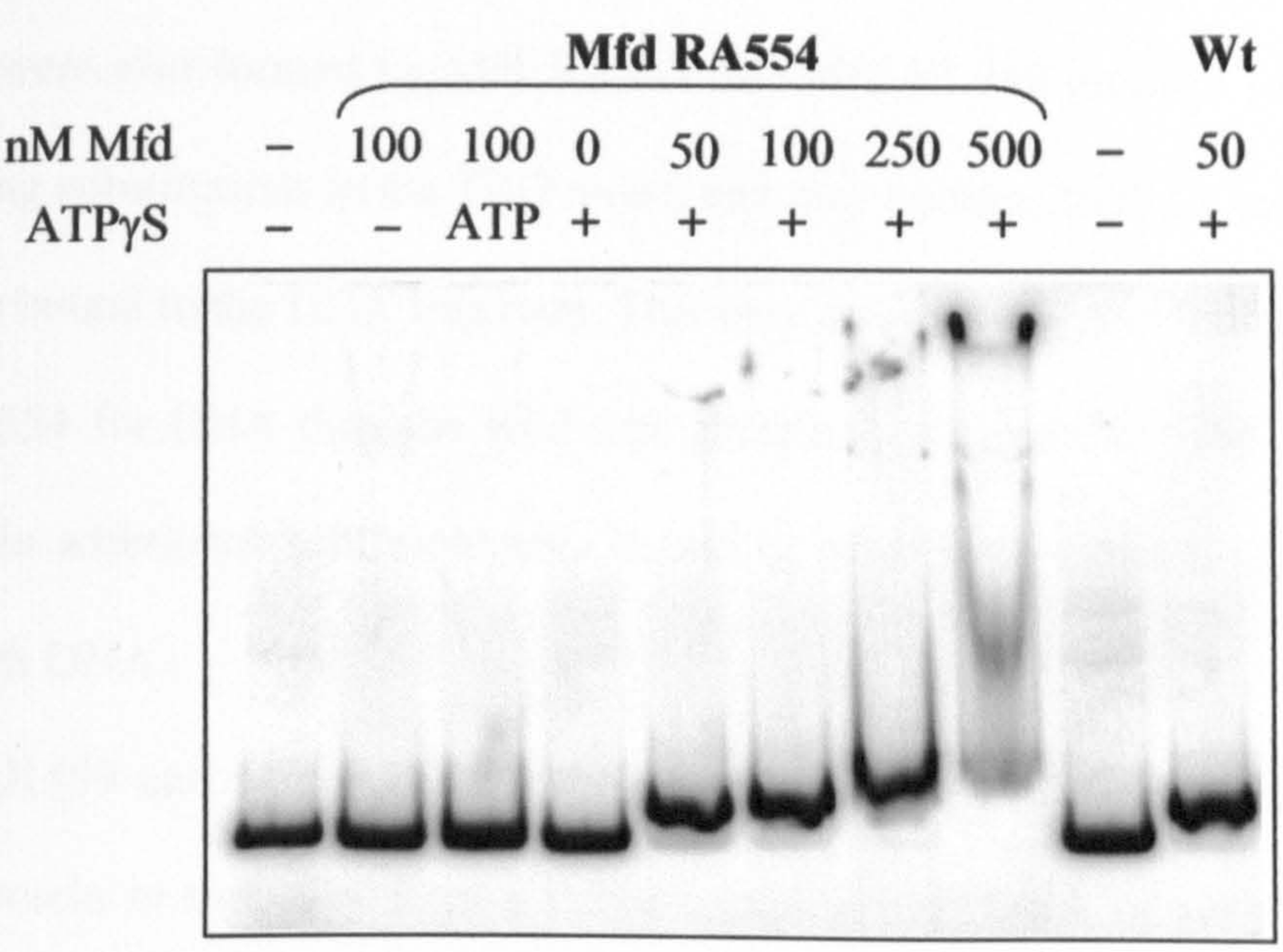


Figure 6.8. DNA binding activity of LR499 and RA554 Mfd proteins.

Purified wild-type Mfd or Mfd carrying the indicated substitutions was incubated for 25 minutes at 37°C with an *Eco*RI-*Bam*HI end-labelled fragment from pSRc33. 2 mM ATP or ATPγS was included as indicated. The complexes were analysed using 5% acrylamide gels.

Wild-type Mfd protein forms stable complexes with DNA in the presence of ATPγS (figure 6.9 A lane 2 and 6.9 B lane 10) but not in the presence of ATP or in the absence of nucleotide (see figure 4.8). Mfd LR499 behaved the same as the wild-type protein in DNA binding assays, also binding DNA in an ATPγS-dependent manner (figure 6.8 A). Similarly, complexes with lower electrophoretic mobility than the DNA fragment alone were observed upon incubation with Mfd RA554 in the presence of ATPγS (figure 6.8 B lanes 5-8) but not in the absence of nucleotide (lane 2) or in the presence of ATP (lane 3). The complexes formed with 250 nM or 500 nM Mfd RA554 in the presence ATPγS possessed lower electrophoretic mobility than the complexes formed at 50 nM (compare lanes 7 and 8 with lane 5), suggesting the formation of higher order complexes. These highly retarded complexes were also formed by Mfd RA953 and Mfd HL948 proteins (see chapter 4), containing substitutions in the TRG motif, and may contain multiple molecules of Mfd protein bound to the DNA fragment. This may indicate a tighter binding affinity of Mfd RA554 for DNA than the wild-type protein. Alternatively the complexes could contain additional Mfd molecules bound to Mfd protein that is already in a complex with DNA.

Both Mfd LR499 and Mfd RA554 possess similar DNA binding properties to the wild-type protein in that they form a stable complex with DNA in the presence of ATPγS but not in the absence of nucleotide or in the presence of ATP and therefore a defect in DNA binding does not appear to be responsible for the defect that these proteins showed in RNAP displacement *in vivo*.

Elongation complex displacement by Mfd proteins containing substitutions in the RID

Contact between Mfd and RNAP is expected to be required both for recruitment of Mfd to the site of the stalled elongation complex and for RNAP displacement. Therefore substitution of residues essential for the interaction between Mfd and RNAP should result in a defect in elongation complex dissociation. When tested *in vivo,* Mfd LR499 failed to increase the efficiency of roadblock repression both on solid media in the *cat* reporter system and in liquid culture using the *luc* reporter, indicating that the protein was defective in elongation complex displacement *in vivo*. Mfd RA554 conferred a chloramphenicol-resistant phenotype on solid media to cells containing the *cat* reporter construct but behaved as wild type in assays performed in liquid culture using the *luc* reporter. To directly assess the ability of purified Mfd LR499 and Mfd RA554 proteins to displace elongation complexes, the purified proteins were tested for their ability to dissociate RNAP *in vitro*.

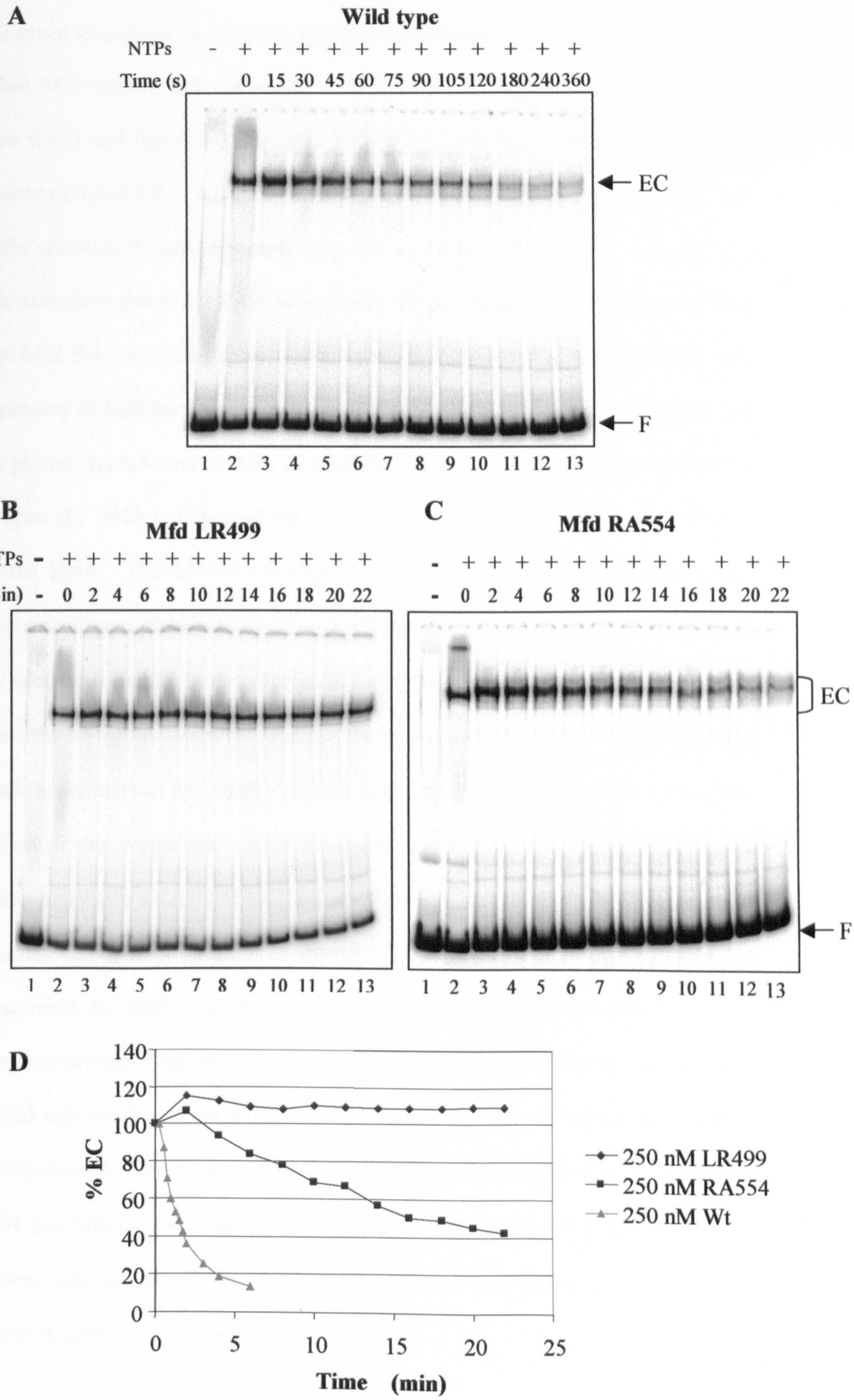
Elongation complexes stalled by nucleotide starvation were formed on the end-labelled *Eco*RI-*Bam*HI fragment of pSRc33. Purified wild-type Mfd, Mfd LR499 or Mfd RA554 protein was added at a final concentration of 250 nM to the stalled elongation complexes. Samples were taken at timed intervals and were analysed by EMSA (figure 6.9 A-C). The amount of elongation complex present at each time point was quantified and expressed as a percentage of the elongation complex that was present before Mfd addition (figure 6.9 D pink squares).

Addition of wild-type Mfd caused rapid dissociation of the stalled elongation complex (figure 6.9 A compare lanes 2 and 13). Approximately 85% of elongation complexes were displaced following 5 minutes incubation with Mfd at 37°C (figure

Figure 6.9. Time course of *in vitro* elongation complex displacement by wild-type Mfd , Mfd LR499 and Mfd RA554.

Initiation complexes were formed by incubation of a ^{32}P end-labelled *EcoR*I-*Bam*HI fragment from pSRc33 with RNAP holoenzyme. Elongation complexes were formed at +33 by the addition of ATP, UTP and GTP (NTPs). (A) Purified wild-type Mfd protein, (B) Mfd LR499 or (C) Mfd RA554 was added at 250 nM. Complexes were analysed using EMSA gels and the bands corresponding to elongation complexes (EC) and the unbound fragment (F) are indicated. Reactions were performed in duplicate and representative gels are shown.

(D)The EMSAs were quantified using ImageQuant software and the percentage elongation complex present at each time point was expressed as a percentage of the elongation complex present at time 0. The graph shows the average of data from two EMSAs for each Mfd protein.


A
Wild type
NTPs - + + + + + + + + + + + +
Time (s) 0 15 30 45 60 75 90 105 120 180 240 360
EC
F
1 2 3 4 5 6 7 8 9 10 11 12 13
B
Mfd LR499
NTPs - + + + + + + + + + + + +
Time (min) - 0 2 4 6 8 10 12 14 16 18 20 22
1 2 3 4 5 6 7 8 9 10 11 12 13
C
Mfd RA554
- + + + + + + + + + + + +
- 0 2 4 6 8 10 12 14 16 18 20 22
EC
F
1 2 3 4 5 6 7 8 9 10 11 12 13
D
% EC
Time (min)
250 nM LR499
250 nM RA554
250 nM Wt

6.9 D green triangles). In contrast, incubation of stalled elongation complexes with purified Mfd LR499 did not result in their displacement even after 22 minutes (figure 6.9 B and figure 6.9 D). Mfd RA554 was still able to displace elongation complexes (figure 6.9 C compare lanes 2 and 13) but did so at a much slower rate than the wild-type protein (figure 6.9 D).

These data show that Mfd LR499 is unable to displace elongation complexes *in vitro* whilst Mfd RA554 retains some displacement activity but at a much-reduced rate. The process of Mfd-mediated elongation complex displacement can be divided into three phases; recruitment of Mfd to stalled elongation complexes, translocation of Mfd upon the DNA upstream of the stalled complex and finally removal of RNAP from the DNA. The LR499 substitution may completely abolish or dramatically reduce recruitment of Mfd to RNAP either by removing one of the contact residues or by causing a change in the conformation of the RNAP interaction domain. The slower rate of displacement by Mfd RA554 in comparison to wild-type Mfd could indicate a decrease in the affinity of Mfd for stalled RNAP rather than a complete abolition of the interaction. Alternatively it may be that the LR499 and RA554 substitutions cause a defect in Mfd translocation or removal of RNAP rather than in the initial recruitment of Mfd to RNAP. If the defect in elongation complex displacement by Mfd RA554 is caused by a reduction in the affinity of the interaction between Mfd and RNAP, recruitment of Mfd would be the rate-limiting step and one would expect the rate of displacement to be increased by a higher concentration of either Mfd RA554 or RNAP. In order to determine whether Mfd RA554 and Mfd LR499 behave as Mfd proteins with a reduced affinity for RNAP, the concentration-dependence of elongation complex displacement by Mfd LR499 and Mfd RA554 was examined.

Firstly, dissociation of RNAP by Mfd LR499 and Mfd RA554 at a single time point was examined at a range of Mfd concentrations. Elongation complexes were stalled as described above on the *Eco*RI-*Bam*HI fragment of pSRc33. Purified Mfd LR499 and Mfd RA554 proteins were added at increasing concentrations and the reactions were incubated for 5 minutes before they were analysed by EMSA (figure 6.10). No decrease in the amount of elongation complex was observed upon incubation with Mfd LR499, even at 480 nM Mfd (compare lanes 1 and 6). In fact some increase in the amount of elongation complex in the gel was observed as the concentration of Mfd LR499 was increased. This suggests that Mfd stabilises the elongation complexes in the gel and this was also observed upon addition of Mfd RA554. Titration of Mfd RA554 into the reaction did not cause elongation complex displacement but instead caused 'supershifting' of the elongation complex (compare lanes 7 and 12). At 30, 60 and 120 nM two bands were observed (lanes 8, 9 and 10), the faster migrating band corresponding to a stalled elongation complex and the band with lower electrophoretic mobility corresponding to the complex that was observed at 480 nM Mfd RA554. This band presumably contains Mfd bound to the elongation complex. Supershifting of elongation complexes by Mfd RA554 was also observed in gels used to analyse the rate of RNAP displacement (compare lanes 2 and 3 in figure 6.9 C and lanes 2 and 3 in figure 6.11). Supershifting of the elongation complex could be due to either Mfd binding to RNAP but failing to displace it or alternatively by Mfd binding to a DNA fragment that is also bound by RNAP. Some binding of Mfd RA554 to the DNA fragment was observed at 250 and 480 nM Mfd protein (figure 6.10 lanes 11 and 12).

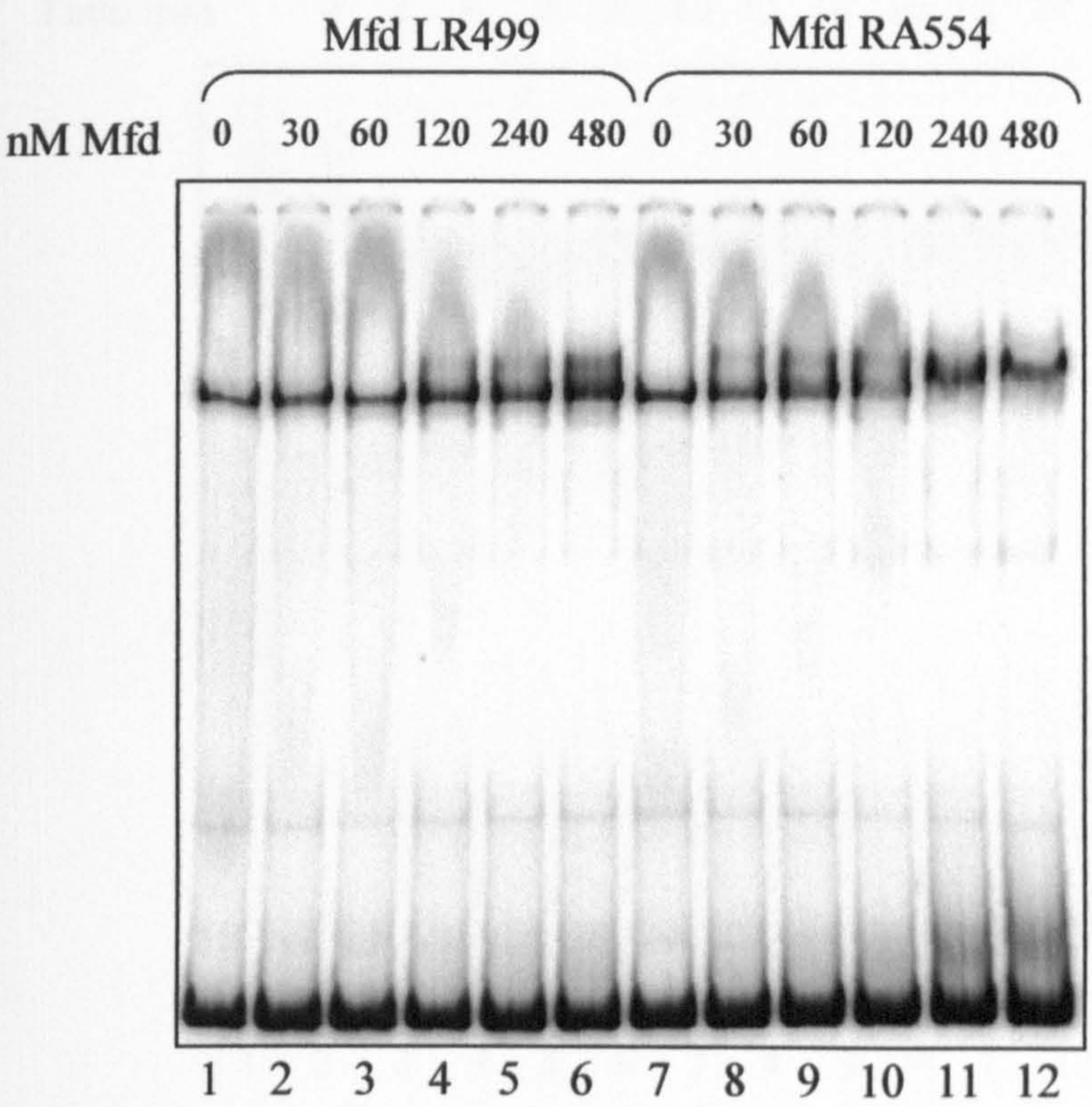


Figure 6.10. Titration of Mfd LR499 and Mfd RA554 into elongation complex displacement assays *in vitro*.

Elongation complexes stalled at +33 were formed by incubation of a ^{32}P end-labelled *EcoR*I-*Bam*HI fragment from pSRc33 with RNAP holoenzyme and ATP, UTP and GTP. Purified Mfd LR499 or Mfd RA554 were incubated with stalled elongation complex for 5 minutes at 37°C before being analysed on a 4.5% acrylamide gel. A representative gel is shown.

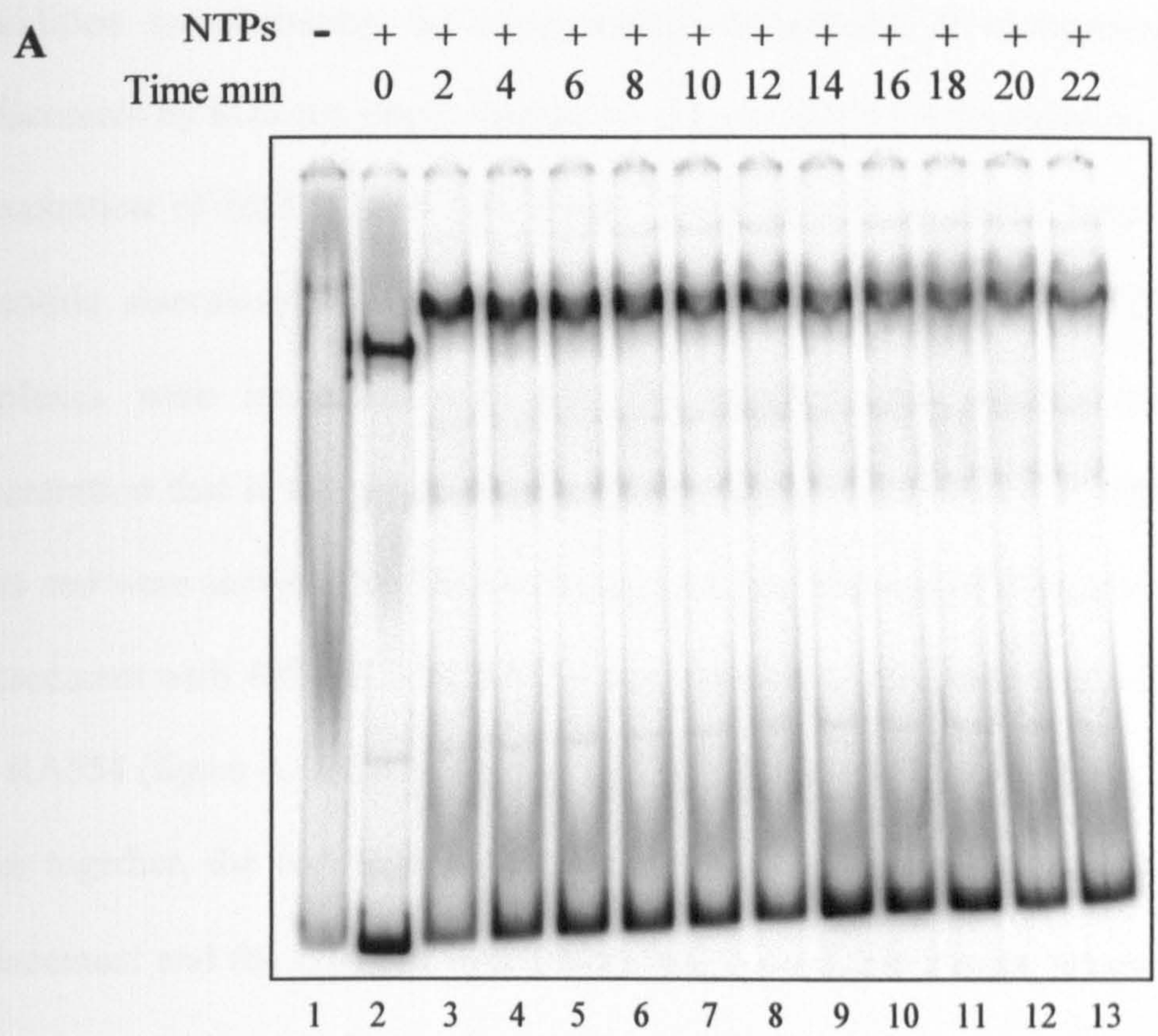


Figure 6.11. The effect of Mfd RA554 concentration on the rate of elongation complex displacement.

(A) Initiation complexes were formed by incubation of a ^{32}P end-labelled *Eco*RI-*Bam*HI fragment from pSRc33 with RNAP holoenzyme. Elongation complexes were formed at +33 by the addition of ATP, UTP and GTP (NTPs). Purified Mfd RA554 was added at a final concentration of 480 nM, samples were taken at different time points and were analysed by EMSA. (B) The amount of elongation complex remaining at each time point was expressed as a percentage of the elongation complex present at time 0. The rate of displacement by 250 nM Mfd RA554 (figure 6.9) has been overlayed for comparison.

In addition to examining the concentration-dependence of elongation complex displacement by Mfd at a single time point, the rate of RNAP dissociation at a higher concentration of Mfd RA554 was tested. Elongation complexes were stalled by nucleotide starvation on the *Eco*RI-*Bam*HI fragment of pSRc33. The stalled complexes were incubated with 480 nM purified Mfd protein (the highest concentration that it was practical to use) and samples were taken at different time points and were analysed by EMSA (figure 6.11 A). The rate of elongation complex displacement with 480 nM Mfd RA554 was identical to that observed with 250 nM Mfd RA554 (figure 6.11 B).

Taken together, the results of experiments examining rates of elongation complex displacement and the effect of Mfd concentration on displacement suggest that Mfd LR499 is unable to dissociate elongation complexes *in vitro*. No displacement of RNAP was observed even following a long incubation with Mfd LR499 or with increased concentrations of Mfd LR499. The simplest explanation of these data is that the LR499 substitution disrupts recruitment to RNAP since this amino acid lies within the RID of Mfd. However, we cannot absolutely rule out the possibility that L499 is required for Mfd translocation or RNAP removal.

The fact that doubling the concentration of Mfd RA554 had no effect on the rate of elongation complex displacement and that Mfd RA554 was able to supershift an elongation complex, suggests that this substitution is not affecting RNAP recruitment directly. The data is more consistent with the rate of displacement by Mfd RA554 being limited by translocation by Mfd or removal of RNAP. This does not eliminate the possibility that R554 forms a contact with RNAP during translocation or removal of RNAP, but suggests that it is not necessary for initial recruitment to the stalled elongation complex. Residues essential for RNAP

displacement that are proposed to affect translocation by Mfd have been identified and are located in a different region of the protein (in the TRG motif and helicase domains, see chapter 4) but it is possible that ATP hydrolysis also needs to be coupled to conformational changes in other domains of the Mfd protein to cause removal of RNAP from the DNA template.

Effect of substitutions in the RID on the interaction with amino acids 1-142 of the RNAP β subunit

Analysis of elongation complex displacement (both *in vivo* and *in vitro*) is an indirect method of testing whether substitutions in the RNAP interaction domain of Mfd affect RNAP recruitment. To examine more directly whether any of the substitutions identified in the RNAP interaction domain, which were deficient in elongation complex displacement, are involved in initial recruitment to RNAP they were tested in the yeast two-hybrid assay. This assay was used to demonstrate an interaction between amino acids 472-603 of Mfd and the first 142 amino acids of the β subunit of RNA polymerase (Park *et al.*, 2002) and it was this interaction that was targeted by site-directed mutagenesis. Therefore the effect of HA485, LR499 or RA554 substitutions on the interaction between these two domains in the yeast two-hybrid assay (figure 6.12 A) was examined.

The fragment of *mfd* encoding amino acids 472-603, either with the wild-type sequence or containing HA485, LR499 or RA554 substitutions was cloned into pAS-2.1 to produce the plasmids pAS-Mfd, pAS-HA485Mfd, pAS-LR499Mfd and pAS-RA554Mfd respectively. Problems were later encountered with the growth of the yeast in liquid media that prevented the analysis of the β-galactosidase activity of transformants containing pAS-HA485Mfd. The yeast strain Y190 was transformed with pACT-RpoB and either pAS2-1 or pAS-LR499Mfd or pAS-RA554Mfd and the

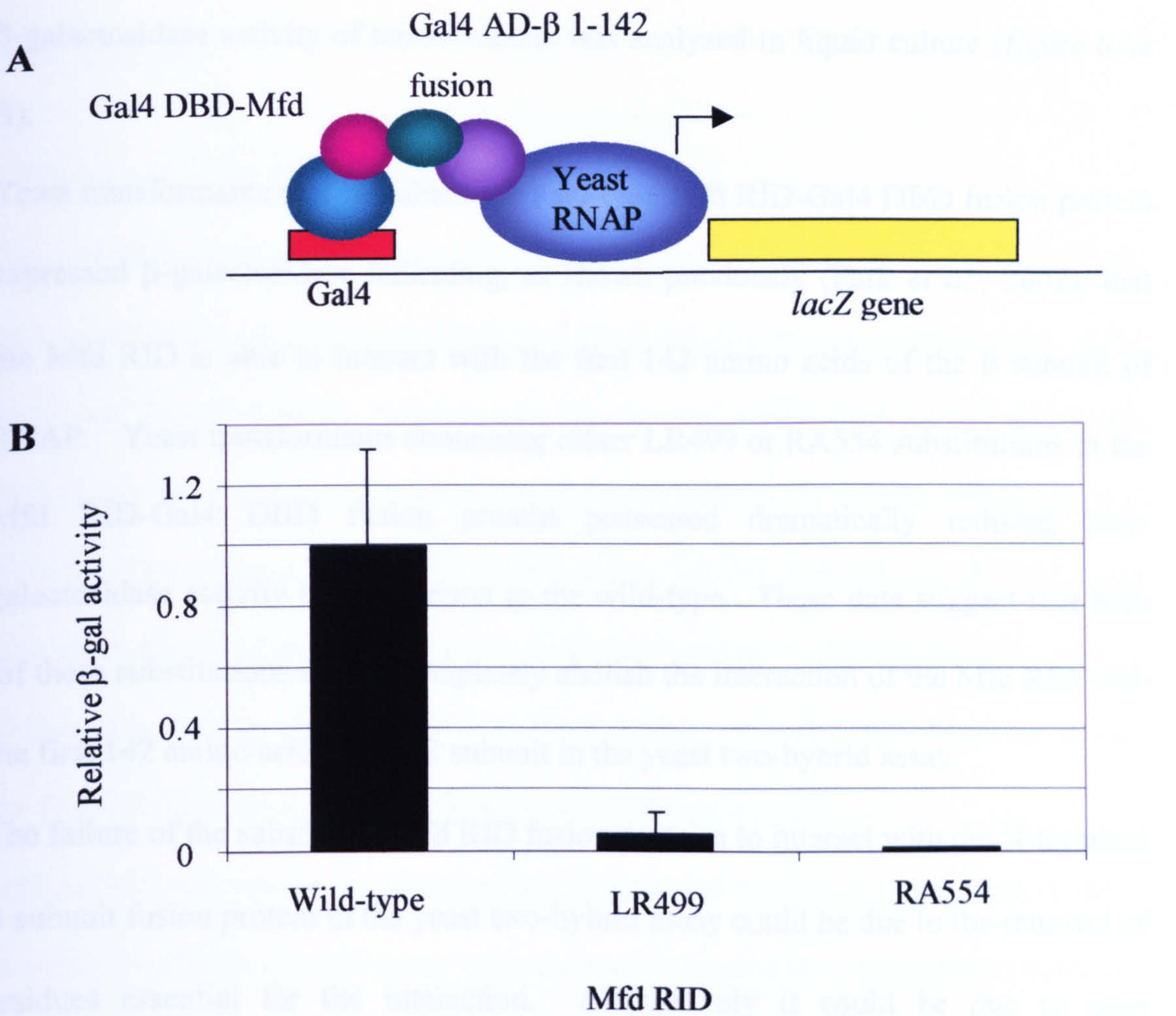


Figure 6.12. Interaction of amino acids 1-142 of the β subunit with the RID of Mfd in yeast two-hybrid assay.

(A)The yeast strain Y190 contains a chromosomal copy of the *lacZ* gene whose expression is controlled by Gal4. Fusion proteins containing the Mfd RID (amino acids 472-603) fused to the C-terminus of the Gal4 DNA-binding domain, and the first 142 amino acids of the β subunit of RNA polymerase fused to the C-terminus of the activation domain of Gal4 are expressed from pAS-Mfd and pACT-RpoB respectively. Interaction of the two fusion proteins will bring together the DNA-binding domain and the activation domain of Gal4, leading to *lacZ* expression.

(B) Competent Y190 yeast were transformed with pACT-RpoB and either pAS-Mfd, pAS-LR499Mfd or pAS-RA554Mfd. β-galactosidase activity was measured from 3 ml overnight culture grown in SD LT^- media. β-galactosidase activity was normalised to the optical density at which the culture was harvested. Activities are expressed relative to the normalised β-galactosidase activity of cells containing the wild-type Mfd RID. The values shown are the average of three independent experiments with s.d.

β-galactosidase activity of transformants was analysed in liquid culture (figure 6.12 B).

Yeast transformants that contained the wild-type Mfd RID-Gal4 DBD fusion protein expressed β-galactosidase indicating, as shown previously (Park *et al.*, 2002), that the Mfd RID is able to interact with the first 142 amino acids of the β subunit of RNAP. Yeast transformants containing either LR499 or RA554 substitutions in the Mfd RID-Gal4 DBD fusion protein possessed dramatically reduced beta-galactosidase activity in comparison to the wild-type. These data suggest that both of these substitutions almost completely abolish the interaction of the Mfd RID with the first 142 amino acids of the β subunit in the yeast two-hybrid assay.

The failure of the substituted Mfd RID fusion proteins to interact with the N-terminal β subunit fusion protein in the yeast two-hybrid assay could be due to the removal of residues essential for the interaction. Alternatively it could be due to poor expression or incorrect folding of the fusion proteins in yeast. Confirmation of the expression levels of the Mfd fusion proteins in yeast would remove the possibility that the lack of interaction between the mutant Mfd RIDs and the β subunit was due to low protein levels.

DISCUSSION

Substitutions in the RNAP interaction domain of Mfd

The RNAP interaction domain has previously been localised to amino acids 472-571 by a combination of pull-down assays and yeast two-hybrid assays (Park *et al.*, 2002; Selby and Sancar, 1995a). Random and site-directed mutagenesis of this region was carried out to try and identify which residues are responsible for the interaction with RNAP. Three single amino acid substitutions, LR499, HA485 and RA554, were identified that resulted in a defect in elongation complex displacement *in vivo*. A further substitution, HL527, was identified, which abolished Mfd function *in vivo* but Mfd HL527 was shown to be poorly expressed or unstable.

Mfd LR499 was unable to increase the efficiency of roadblock repression in the *in vivo luc* roadblock reporter assay and was unable to displace elongation complexes *in vitro*. The LR499 substitution also abolished the interaction of the Mfd RID with the N-terminus of the β subunit of RNA polymerase in the yeast two-hybrid assay. These observations indicate that the LR499 substitution abolishes (or dramatically reduces) the ability of Mfd to be recruited to RNAP. The effect could be direct or alternatively indirect due to the substitution causing misfolding of the RID, or a change in the conformation or local charge in the vicinity of a residue that contacts RNAP. While two of the residues identified in the Mfd contact patch of RNAP were charged (K118 and E119), the third residue of the patch was a small aliphatic amino acid (I117). Therefore it could be that L499 in Mfd is the binding partner of I117 in the β subunit of RNAP.

Mfd HA485 conferred chloramphenicol-resistance on solid media in the *cat* roadblock reporter system and showed higher luciferase activity than the wild-type

protein in the *luc* reporter assay. Although the HA485 substitution caused a defect in elongation complex displacement in the *luc* reporter assay when this substitution was combined with a DA604 substitution (a substitution next to the Q-tip that appeared to result in a defect in ATP hydrolysis but that was still able to displace elongation complexes both *in vitro* and *in vivo,* see chapter 5) the protein was once again able to function *in vivo*.

There are a number of possible explanations for these observations. Firstly the substituted proteins could have different levels of expression or the HA485 Mfd protein may be less stable than Mfd containing both HA485 and DA604 substitutions. Alternatively, D604 may have an inhibitory effect on RNAP displacement, which is silenced by interaction of this amino acid with H485 or with a region of the Mfd protein that is held in the correct conformation by H485. If this were true the inhibitory effect of D604 would be silenced in the wild-type protein. However, in HA485 substituted proteins the inhibitory effect of D604 would become apparent due to the absence of a compensatory charge and in the double mutant the inhibitory effect would be removed. The inhibitory effect could either be structural, due to misfolding of the mutant protein or functional e.g. the presence of an un-neutralised negatively charged amino acid affecting RNAP recruitment, DNA binding, ATP binding or ATP hydrolysis.

One specialised example of this model is that H485 could play a role in coupling recruitment of Mfd to RNAP to the ATP hydrolysis domain. Binding of Mfd to an elongation complex could result in stimulation of the weak ATPase activity located within the helicase domains in order to cause elongation complex displacement. Given that substitution of D604, which is located next to the ATP binding pocket,

alters the effect of an H485 substitution located within the RNAP interaction domain it may be that there is communication or coupling between these two domains. Although Mfd HA485, containing a single substitution, was expressed, it was not possible to purify the protein using the three-column affinity chromatography protocol that had been used to purify the wild-type protein and other Mfd mutants. In order to study the properties of this protein *in vitro* modification of the native protein purification protocol could be attempted or the protein could be His-tagged and purified using a nickel column. The fact that HA485 did not purify by the same method as the wild-type protein may favour the suggestion that the protein is misfolded and that H485 is a structurally important residue. Both H485 and D604 are 100% conserved and this suggests these residues are important for Mfd function. Mfd RA554 behaved as wild type in the *luc* reporter assay in liquid culture but was chloramphenicol-resistant when examined with the *cat* system, reflecting the greater sensitivity of the roadblock repression assay on solid media. Purified Mfd RA554 was still able to displace elongation complexes *in vitro* but at a much-reduced rate in comparison to the wild-type protein. The rate of RNAP dissociation by Mfd RA554 was shown to be concentration-independent and Mfd RA554 was able to supershift stalled elongation complexes. Both Mfd LR499 and Mfd RA554 were still able to hydrolyse ATP and bind DNA in the same ATPγS-dependent manner as the wild-type protein. These experiments suggest that the RA554 substitution does not cause a defect in recruitment of Mfd to RNAP but instead causes a defect in a later step in displacement such as Mfd translocation or RNAP removal. Surprisingly, when placed into the yeast two-hybrid assay, the RA554 Mfd RID fusion protein failed to interact with the 1-142 β subunit fusion protein. This could indicate that R554 is actually involved in the interaction between these two domains or that its

substitution decreases the expression of the fusion protein or causes incorrect folding of the RID domain.

In a homology model of the helicase domains of Mfd, R554 is located in a long α-helix, which links the RNAP interaction domain of Mfd to the helicase domains (figure 6.13). The position of R554 is close to where the short helix that follows the loop of the TRG motif crosses this long helix and it is possible that ionic interactions occur between R554 and two glutamate residues present in the short helix (E977 and E980) located approximately 6.8 Å away. This is a slightly longer distance that the 4-5 Å that is usually defined as the limit for salt-bridge formation, but only a slight reorientation of one of the side chains would be required to bring these residues within this distance. In addition, dipole interactions could occur over this distance even in the absence of a salt-bridge. Loss of the positively charged R554 could therefore affect the conformation of the TRG motif which is located immediately C-terminal to the short crossing helix. Alternatively, substitution of R554 could alter the conformation of the long helix and consequently the position of the RNAP interaction domain relative to the other domains of Mfd. Either of these could result in a defect in elongation complex displacement.

A conformational change in the Mfd protein could be required during the process of RNAP displacement, either during translocation or RNAP removal, and it could be that R554 is required for this change or is required for a step following the change, for example it could form a contact with RNAP that is not required for initial recruitment but that is required for displacement. To further study the effect of LR499 and RA554 substitutions on RNAP binding, more direct assays of interaction with RNAP rather than RNAP displacement could be employed such as pull-down assays, crosslinking or native gel electrophoresis.

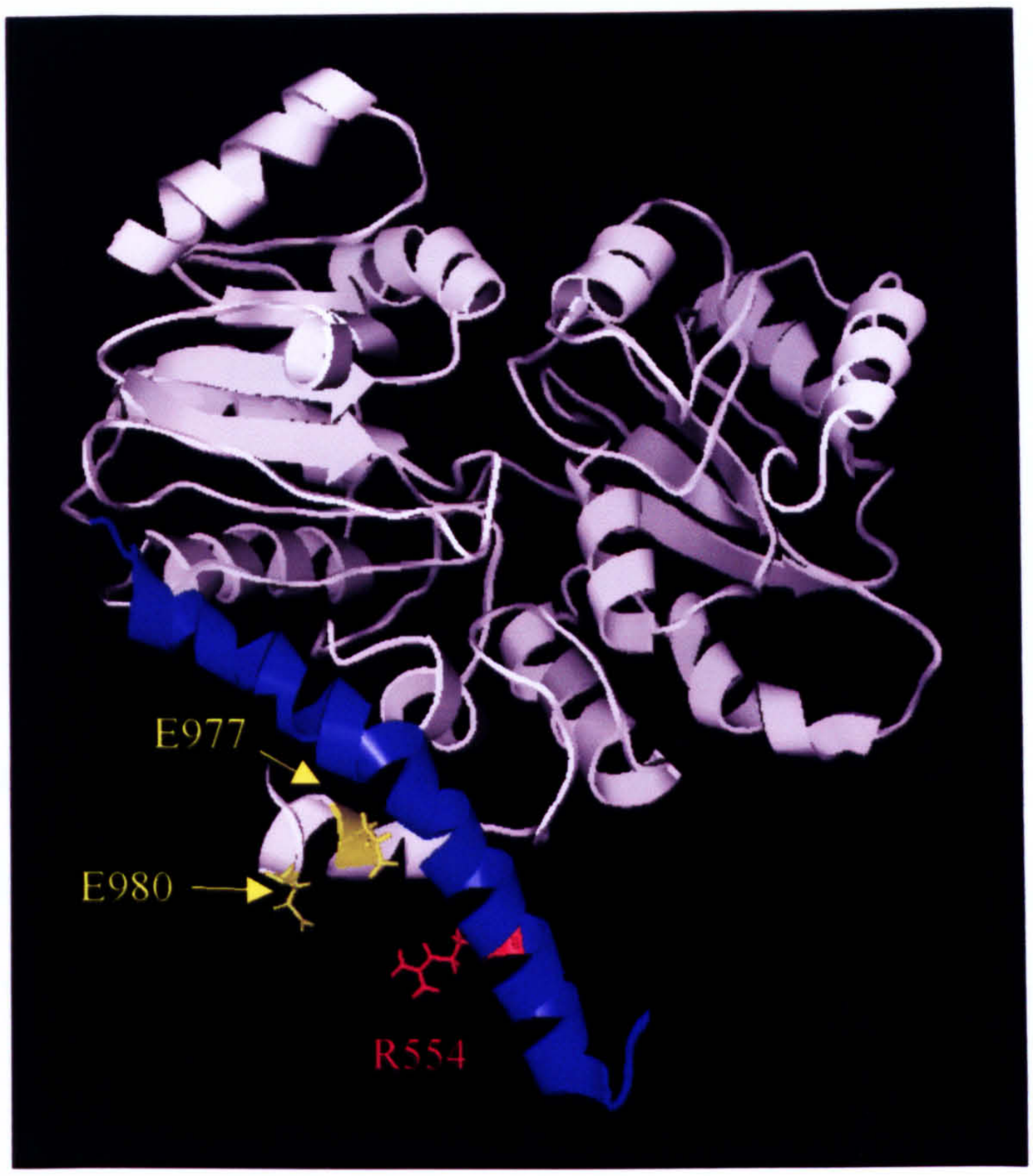


Figure 6.13. Location of R554 in the homology model of the helicase domains of Mfd.

R554 is shown in red on the model of the helicase domains, produced by homology modelling onto the crystal structure of RecG. The long helix which links the RNAP interaction domain and the helicase domains is shown in blue and was included in both the random mutagenesis screening and in site-directed mutagenesis studies. The remainder of the RID is not shown. R554 is located in the long helix close to the where a short helix that follows the TRG motif crosses. Two negatively charged glutamate residues (E977 and E980, shown in yellow) are located in this short helix and could form ionic interactions with R554.

All of the charged amino acids in the RID of *E. coli* Mfd that are conserved with *B. subtilis* were targeted, and many of the substituted proteins remained able to increase the efficiency of roadblock repression. Therefore many of the charged residues in the RNAP interaction domain of Mfd are not essential for RNAP recruitment or displacement. The possibility remains that they could contribute to an RNAP contact site formed by several residues and although their substitution weakens the affinity of the interaction it does not do so to an extent that will abolish binding at the protein concentrations present in the *in vivo* assay.

Given that the Mfd protein from *B. subtilis* is able to displace stalled *E. coli* RNAP substitution of amino acids in *B. subtilis* Mfd equivalent to those identified as important in the RNAP interaction domain of *E. coli* Mfd (H485, L499 and R554 in *E. coli* are equivalent to H508, L522 and R575 in *B. subtilis*) would be expected to have similar effects.

The RNAP interaction domains of *E. coli* Mfd and *B. subtilis* Mfd share 25% and 26% identity with the N-terminal domain of CarD, a transcription factor from *M. xanthus* (Padmanabhan *et al.*, 2001). It is conceivable that this domain acts as an RNAP-binding module in both of these proteins, perhaps forming a similar fold. There is no CarD homologue present in *E. coli* so comparison of conserved residues that may specifically contact *E. coli* RNAP is not possible.

In order to provide confirmation of the critical residues involved in the interaction, charge reversal mutants could be synthesised in the β subunit of RNAP and the RID of Mfd. The single substitutions examined would be expected to result in an absence of stalled elongation complex displacement but this activity may be restored when charge reversal substitutions are made in both Mfd and the β subunit. Clearly a co-

crystal structure of Mfd or even the RID of Mfd bound to core RNAP would help to elucidate the details of the interface between these two proteins further.

Screening procedures

From the screen of the RID of Mfd for mutants that are defective in RNAP displacement, isolation of single amino acid substitutions was rare. This could be the result of a number of factors. The majority of chloramphenicol-resistant clones from initial screens, contained plasmids with large-scale deletions. Since the screen identified loss-of-function mutations, any plasmid that contains a disrupted *lacI*^{q} gene would be chloramphenicol-resistant. To try and minimise this problem a secondary screen was introduced to eliminate any clones lacking a functional *lacI*^{q} gene. This reduced the number of plasmids containing large-scale deletions in the latter stages of the screening procedure but the frequency of isolating single amino acid substitutions remained low.

Using too many error-prone PCR cycles to increase the mutation frequency was a problem since it resulted in the generation of multiple substitutions. Therefore there is a fine balance between a sufficiently high mutation frequency to isolate substitutions above the background of plasmid deletions and a low enough mutation frequency to isolate only single or double substitutions. Additionally there is no guarantee that a single substitution will be sufficient to abolish the interaction between Mfd and RNAP. Several amino acids may well contribute to the binding patch and elimination of any one of these residues in isolation may be insufficient to abolish the effect of the Mfd protein in roadblock repression, since the other residues in the contact patch would remain intact.

CHAPTER 7

DISCUSSION

Single amino acid substitutions that abolish the ability of Mfd to displace stalled elongation complexes *in vivo* and *in vitro* have been identified in three regions of the Mfd protein. The first of these regions is the RNAP interaction domain and the substitutions located in this domain may cause defects in displacement due to an inability to contact RNA polymerase and be recruited to the elongation complex or due to defects in the process of RNAP removal from the DNA template. Secondly, a substitution in the Q-tip region of Mfd resulted in a defect in nucleotide binding and therefore an inability to hydrolyse ATP or CTP in order to drive elongation complex displacement. Finally substitutions were identified in a region of Mfd downstream of its seven superfamily 2 helicase motifs. This region is homologous to the TRG motif of RecG, shown to be required for coupling ATP hydrolysis to translocation on dsDNA. The requirement of an intact TRG motif for Mfd-mediated displacement of elongation complexes supports a mechanism for termination of transcription by Mfd that involves translocation of Mfd on upstream DNA. In addition a novel *in vivo* assay for Mfd function has been developed, which is dependent on the ability of Mfd to displace an elongation complex stalled at a protein roadblock. The assay provides a readily monitored phenotype for *mfd*$^{+}$ cells and a means of screening libraries for *mfd* mutants defective in elongation complex displacement. Finally it was shown that Mfd is able to bind to 32 bp of dsDNA, but does so with lower affinity than to longer (250 bp) fragments of DNA; this suggests that the Mfd protein may be able to wrap DNA. Identification of amino acids essential for Mfd-mediated elongation complex displacement and studies of the properties of these Mfd proteins permit development of a model for termination by Mfd.

The TRG motif has been proposed to be involved in coupling ATP binding and hydrolysis by the seven traditional helicase motifs to translocation upon dsDNA

(Mahdi *et al.*, 2003). The involvement of the TRG motif of Mfd in elongation complex displacement suggests that Mfd may translocate along DNA in the same manner as RecG and that the RNAP interaction domain attached to the motor domain couples translocation to RNAP dissociation. DNA upstream of a stalled elongation complex has been shown to be essential for Mfd-mediated displacement and Mfd is also able to restore arrested backtracked complexes to productive elongation (Park *et al.*, 2002), suggesting that Mfd acts as a forward translocase. This could be viewed as either pulling the DNA out of the back of the elongation complex or as pushing the elongation complex forward on the DNA template.

Forward translocation of RNA polymerase without accompanying RNA synthesis, caused by nucleotide starvation or an inability to incorporate nucleotides opposite a lesion, would result in shortening of the DNA:RNA hybrid. It has been shown that reducing the length of the hybrid from 8-9 bp to 5 bp decreases the stability of an elongation complex (Komissarova *et al.*, 2002). Therefore the result of Mfd forward translocation and rewinding of upstream DNA may be to shorten the hybrid, causing transcription bubble collapse, elongation complex destabilisation and resulting in removal of RNAP from the template. At each position on a transcription template termination competes with incorporation of the next nucleotide, with nucleotide incorporation being favoured under elongation conditions. The inability to incorporate the next nucleotide opposite a lesion in the DNA or under conditions of nucleotide starvation would increase the probability of termination relative to elongation.

It has been proposed that forward translocation of RNAP is also required for intrinsic termination (Santangelo and Roberts, 2004). Hairpin folding was observed to cause shortening of the hybrid from 8 bp to 5 bp (Komissarova *et al.*, 2002). It has been

shown that rewinding of upstream DNA and unwinding of the hybrid is necessary for termination induced by oligonucleotides that mimic the effect of hairpin folding in a natural terminator (Nudler *et al.*, 1995; Ryder and Roberts, 2003; Santangelo and Roberts, 2004). Intrinsic termination was inhibited when blocks to RNAP forward translocation, such as protein roadblocks or psoralen crosslinks that prevent downstream DNA unwinding, were placed downstream of the termination site (Santangelo and Roberts 2004). It could be that in cases where translocation of the transcription bubble is impeded that termination occurs by an alternative mechanism, for example movement of RNAP forward on duplex DNA. It is also possible that Rho-dependent termination involves shortening of the hybrid by Rho 'pulling' the RNA from the elongation complex, which could also push RNAP forward on the template.

As for backtracked complexes, Mfd-meditated translocation of RNAP in the direction of transcription would account for the realignment of the 3′ end of the RNA with the active site and rescue of these arrested complexes. The relative importance of GreB and Mfd in rescue of backtracked elongation complexes *in vivo* is unknown. Mfd was able to displace elongation complexes stalled at roadblocks or lesions as well as by nucleotide starvation. ECs arrested at a roadblock could have backtracked or alternatively could be right up against the protein roadblock. Termination by Mfd has been shown to occur from the 'forward' position i.e. with the 3′ end of the RNA in the active site of RNAP. Once backtracked ECs have been restored to the forward position where movement forward is physically blocked by the protein roadblock, further forward translocation of RNAP could result in the displacement of the roadblock or the roadblock protein could remain bound. If the roadblock remains bound the action of Mfd could result in 'scrunching' i.e. the RNA polymerase rear

end being pushed closer to the front edge resulting in a shorter RNA polymerase footprint and a decrease in length of the transcription bubble. If the roadblock is displaced by RNAP forward translocation, one might expect transcription to continue if the only role of Mfd in EC displacement is its ability to push RNAP forwards. Cooperation between multiple RNA polymerase molecules transcribing the same template was observed to reduce pausing and arrest and to permit read-through of protein roadblocks (Epshtein and Nudler, 2003; Epshtein *et al.*, 2003). Since the action of simply pushing forward a stalled elongation complex at a roadblock does not in itself cause termination, it implies displacement of RNAP by the Mfd protein is not simply a result of forward translocation.

Under conditions in which transcription can proceed i.e. when nucleotides are available and the passage of RNAP is not blocked, Mfd is able to return backtracked elongation complexes to productive elongation. However, when transcription is blocked by a protein roadblock or a thymine dimer, or by nucleotide starvation, the result of Mfd action is dissociation of the stalled complex and therefore termination of transcription (figure 7.1). As an aside, it was noted that Mfd appeared to increase end-to-end transfer of ECs during transcription, generating transcripts longer than the end of the fragment (Selby and Sancar, 1993a). Although this may not be physiologically relevant, since there are no "ends" to the bacterial chromosome, understanding the mechanism of this process may provide clues to both the mechanism of Mfd action on transcription complexes and also on the mechanism of transcription itself such as the ability of elongation complexes to bypass nicks and gaps in the template strand. One occasion when 'ends' are present in the bacterial genome is following exposure to ionising radiation when double-strand breaks

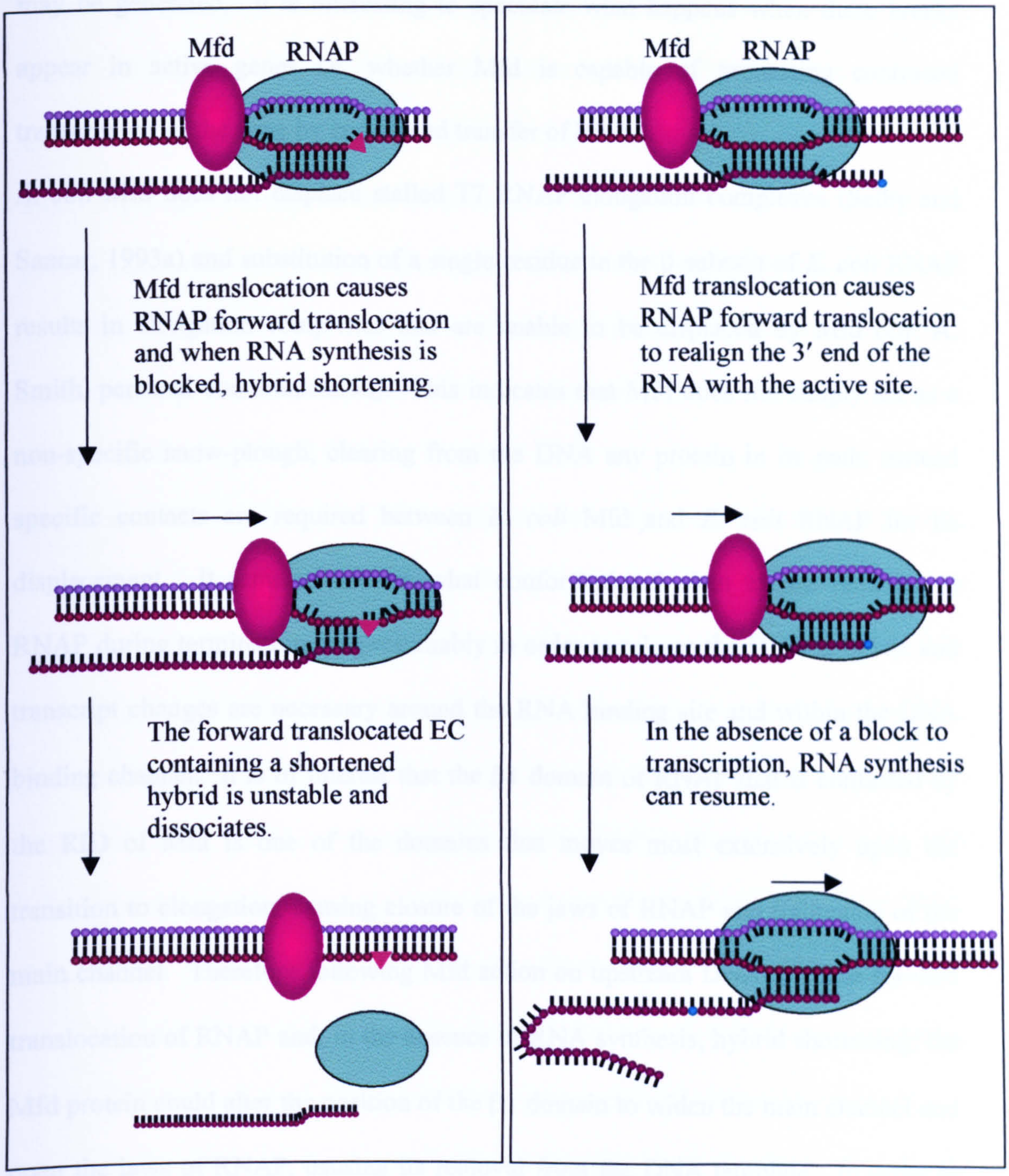


Figure 7.1. Model for Mfd action in the presence and absence of blocks to transcription.

(A) Coupling of ATP hydrolysis to DNA translocation via the TRG motif of Mfd causes forward translocation of RNAP. When transcription is blocked either by nucleotide starvation or by the presence of a DNA-bound protein or a thymine dimer, the Mfd-mediated forward translocation of RNAP results in termination. (B) When further transcription is not blocked for example in a complex that has simply backtracked and become arrested, forward translocation of RNAP allows resumption of RNA synthesis.

may be generated. It is interesting to speculate what happens when these breaks appear in active genes i.e. whether Mfd is capable of promoting continued transcription of the gene by end-to-end transfer of RNAP.

E. coli Mfd does not displace stalled T7 RNAP elongation complexes (Selby and Sancar, 1993a) and substitution of a single residue in the β subunit of *E. coli* RNAP results in elongation complexes that are unable to be displaced by Mfd (Dr. A. Smith, personal communication). This indicates that Mfd does not simply act as a non-specific snow-plough, clearing from the DNA any protein in its path; instead specific contacts are required between *E. coli* Mfd and *E. coli* RNAP for its displacement. It remains unclear what conformational changes are required in RNAP during termination but presumably in order to release the DNA template and transcript changes are necessary around the RNA binding site and within the DNA binding channel. It is of interest that the β1 domain of RNAP that is contacted by the RID of Mfd is one of the domains that moves most extensively upon the transition to elongation, causing closure of the jaws of RNAP and tightening of the main channel. Therefore following Mfd action on upstream DNA to cause forward translocation of RNAP and, in the absence of RNA synthesis, hybrid shortening, the Mfd protein could alter the position of the β1 domain to widen the main channel and open the jaws of RNAP, causing its removal from the DNA template. In order to test whether a specific contact between Mfd and the β1 domain of RNAP is required for displacement, the Mfd protein/motor domains could be fused to another protein that contacts a region/subunit of RNAP on the upstream DNA side. If the function of Mfd in displacement is simply to cause forward translocation of RNAP, then attachment of the motor domain of Mfd to the back end of polymerase would be expected to cause termination, however if a specific contact with the β1 domain is

necessary transcription termination would not occur. Alternatively the RID of Mfd could be fused to the motor domains of another helicase to determine whether the specific contact between the RID and the β subunit restores displacement. From initial experiments it appears that substitution of RA554 in the Mfd RID results in a protein that is defective in removal of RNAP rather than in initial recruitment of Mfd to stalled ECs. This protein could be tested for its ability to rescue backtracked complexes, a process that requires the translocation activity of Mfd, but that does not involve removal of RNAP from the template.

How does Mfd translocate on DNA and how does the action of Mfd on upstream DNA cause forward translocation of RNAP? The SF1 helicase PcrA has been proposed to translocate by changes in affinity of two sites for ssDNA upon ATP binding and hydrolysis and involves flipping of bases into a series of hydrophobic pockets (Velankar *et al.*, 1999). In contrast the SF2 helicase NPHII was shown to translocate using the ribose phosphate backbone of the loading strand. Abasic sites were tolerated in this strand but nicks prevented NPHII motion (Kawaoka, 2004). Several other SF2 helicase proteins including RecG (Mahdi *et al.*, 2003), the Type1 restriction enzymes (Firman and Szczelkun, 2000) and chromatin remodelling factors (Whitehouse *et al.*, 2003) have been proposed to act as double-stranded DNA translocases, contacting the duplex backbone to move along the DNA, negating the requirement for strand separation that occurs in SF1 helicases which primarily contact the DNA bases. It seems likely that Mfd translocates on dsDNA by contacting the phosphodiester backbone.

Mfd on its own or in solution with free DNA or free holoenzyme displays a relatively slow rate of ATP hydrolysis (3-20 min^{-1}) (Selby and Sancar, 1993a), compared to the DNA-stimulated rate of 185 min^{-1} for RecG. This slow rate could

be representative of the rate *in vivo* during translocation and RNAP displacement and may indicate that due to competition between the rate of displacement and the length of the stall, slow ATP hydrolysis by Mfd acts to prevent inappropriate displacement of elongation complexes that are merely paused rather than stalled. Alternatively, hydrolysis of one molecule of ATP could cause translocation of several bp of DNA or relatively few steps forward may be required to cause RNAP displacement; reduction of the length of the hybrid to 5 bp was sufficient to destabilise an elongation complex, equivalent to movement forward of only 4 bp (Komissarova *et al.*, 2002). On the other hand there could be a factor, as yet unidentified, which stimulates ATP hydrolysis by Mfd *in vivo*. Free core RNAP and DNA did not stimulate ATP hydrolysis by Mfd *in vitro* but binding of Mfd to either a stalled elongation complex or to UvrA could trigger a conformational change in the Mfd protein that results in an increase in the rate of ATP hydrolysis. The number of bp translocated for each ATP that is hydrolysed by Mfd has not been determined and indeed translocation itself has not been demonstrated directly. Traditional methods of measuring translocation rates cannot be applied to the Mfd protein since it is unable to displace a triplex-forming oligonucleotide (unpublished observations, cited in (Chambers *et al.*, 2003)), it does not display strand-displacement activity and it does not begin translocation from a specific site on the DNA template. Nor have the conformational states that are associated with different stages of ATP hydrolysis been studied and these could be probed using different ATP analogues to mimic different points in the ATP hydrolysis cycle and examine their effects on DNA binding.

One can envisage translocation of Mfd along dsDNA occurring by one of two mechanisms (figure 7.2). Firstly the Mfd protein could possess two DNA binding

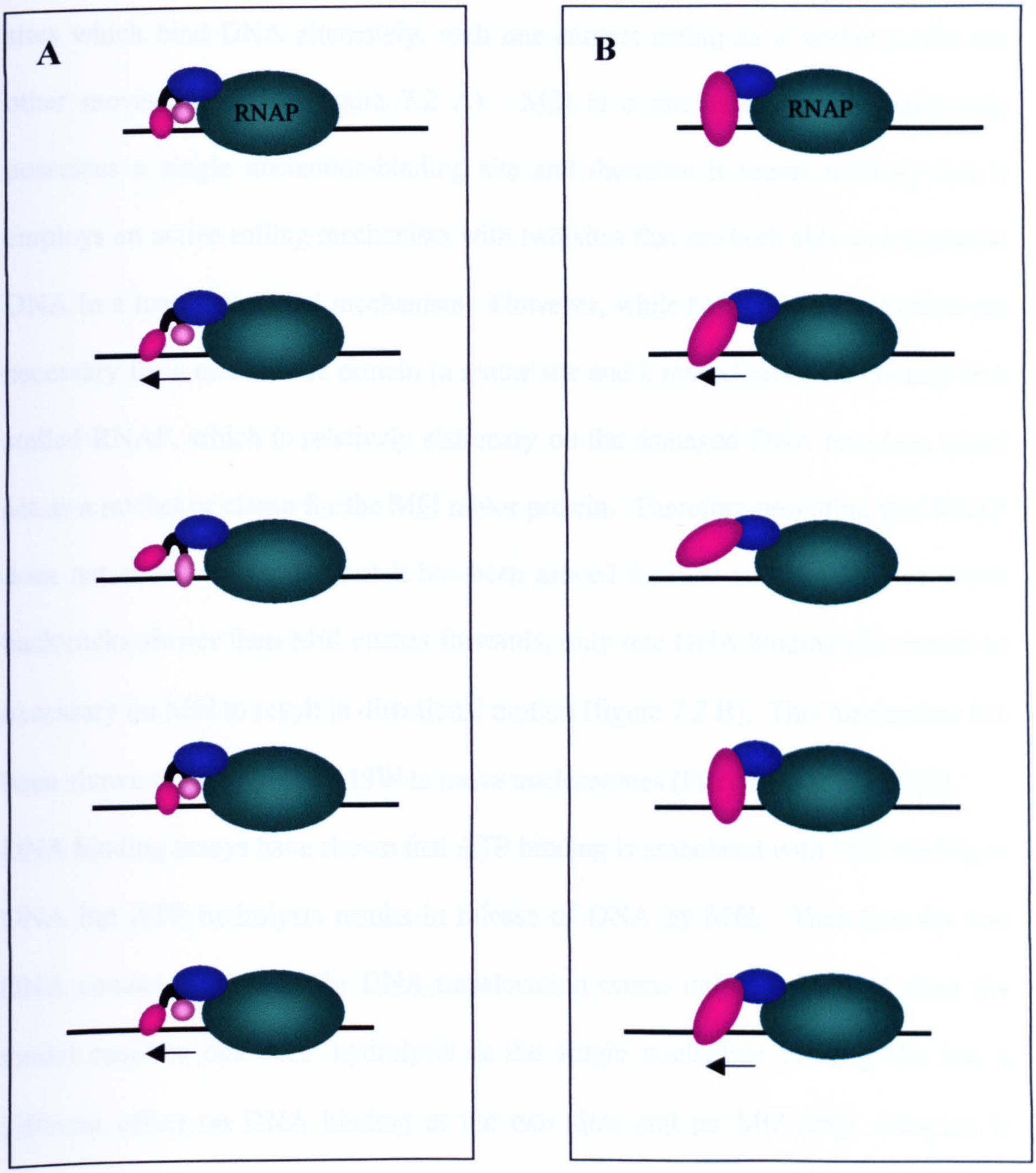


Figure 7.2. Models for DNA translocation by Mfd to induce RNAP forward translocation.

The Mfd protein could either possess two DNA contact sites, one of which acts as a ratchet and the other to move the DNA (A), or the stalled RNAP could effectively act as a ratchet or clamp and the Mfd protein could possess a single DNA contact which undergoes cycles of DNA binding, DNA translocation and DNA release (B).

sites which bind DNA alternately, with one contact acting as a ratchet whilst the other moves the DNA (figure 7.2 A). Mfd is a monomeric protein and only possesses a single nucleotide-binding site and therefore it seems unlikely that it employs an active rolling mechanism with two sites that are both able to translocate DNA in a hand-over-hand mechanism. However, while two DNA contact sites are necessary for a monomeric protein (a motor site and a ratchet site), the contact with stalled RNAP, which is relatively stationary on the damaged DNA template, could act as a ratchet or clamp for the Mfd motor protein. Therefore providing that RNAP does not slide backwards once it has been moved forward on the DNA, or that it backtracks slower than Mfd pushes forwards, only one DNA binding site would be necessary on Mfd to result in directional motion (figure 7.2 B). This mechanism has been shown to be utilised by ISW to move nucleosomes (Fitzgerald *et al.*, 2004).

DNA binding assays have shown that ATP binding is associated with Mfd binding to DNA but ATP hydrolysis results in release of DNA by Mfd. Therefore the two DNA contact site model for DNA translocation seems unlikely for Mfd since the model requires that ATP hydrolysis at the single nucleotide binding site has a different effect on DNA binding at the two sites and no Mfd-DNA complex is observed under conditions when ATP hydrolysis occurs. In order for the two contact model to apply, one of the DNA binding sites would have to be low affinity and insufficient to form a stable Mfd-DNA complex in the absence of an elongation complex, but the affinity of this interaction could be increased by binding to a stalled elongation complex, perhaps by effectively increasing the local concentration of DNA.

Mfd translocation on upstream DNA could result in RNAP forward motion either by pushing RNAP through the contact with the Mfd protein or by the introduction of

torsion into the upstream DNA. Torsion could be introduced to the region of DNA between RNAP and Mfd by linear movement of Mfd along the backbone of the DNA helix. If the single contact model for DNA translocation applies, RNAP would need to move forwards before Mfd releases the DNA or the torsion that has been introduced would be lost. The properties of the upstream DNA that are necessary for Mfd-mediated EC displacement can be tested by reconstituting elongation complexes *in vitro* using oligonucleotides containing modifications such as gaps, abasic sites and psoralen crosslinks. These may allow the directionality of the Mfd translocase to be established, determination of whether Mfd contacts the DNA bases or backbone in order to translocate and whether translocation by Mfd is transmitted to RNAP forward movement through the intervening DNA.

Mot1 and release factor 2 have also been proposed to use helicase action on upstream DNA to displace protein complexes and clues to the mechanism of action of elongation complex displacement by Mfd may be gleaned from studies of these proteins (Darst *et al.*, 2001; Xie and Price, 1997). Mot1 removes TBP (TATA box binding protein), required for recognition of the TATA box found in most eukaryotic promoters, from the DNA to regulate transcription initiation both positively and negatively, depending on the promoter. Mot1 binds to TBP-DNA complexes in the absence of ATP, in contrast to the requirement for ATP for DNA-binding by Mfd, and binding activates the ATP hydrolysis activity of Mot1, contained within its SWI/SNF2 family helicase motifs, to cause dissociation of both TBP and Mot1 (Auble *et al.*, 1997). Footprinting have shown that Mot1 binds to DNA upstream of TBP-DNA complexes and requires 17 bp of upstream DNA to function. There is no evidence that Mot1 translocates along DNA, at least over long distances, although some local tracking is not ruled out (Auble and Steggerda, 1999). Bending or

twisting of the upstream DNA was shown not to be required for TBP displacement by introduction of single bp gaps in either strand and furthermore strand separation was shown not to be necessary for TBP displacement by introduction of psoralen crosslinks (Darst *et al.*, 2001). It was proposed that Mot1 may act as a 'plough' either remaining bound to the upstream DNA and an extension pushing off TBP following ATP hydrolysis, or alternatively short range DNA translocation causing TBP to be barged off the DNA.

Drosophila factor 2 is a 154 kDa protein that causes the release of RNA polymerase II transcripts in an ATP (or dATP)-dependent manner. Factor 2 is likely to cause termination through action on the upstream DNA template since its SWI1/SNF2 family ATPase activity is strongly stimulated by dsDNA (Xie and Price, 1997) and the protein is able to cause release of transcripts from ECs that have only transcribed 10 bases and therefore possess RNA protected by the RNAP. Factor 2 is also able to cause transcript release from arrested transcription complexes. Mammalian factor 2 has been demonstrated to disrupt RNA polymerase II or RNA polymerase I ECs stalled by nucleotide starvation or thymine dimers (Hara *et al.*, 1999), leading to the suggestion that it functions as a TRCF.

The vaccinia virus protein NPH-1 is a superfamily 2 helicase protein that causes termination of vaccinia virus RNA polymerase at specific signals (Deng and Shuman, 1998). The termination signal in the RNA transcript is recognised by another factor called VTF, which transmits the signal to terminate transcription to NPH-1. The ATPase activity of NPH-I is stimulated by DNA and in the absence of a termination signal the protein can act as an elongation factor, promoting transcription. Therefore action on upstream DNA appears to be used by a number of proteins that contain helicase motifs to disrupt specific DNA-protein complexes.

Before Mfd translocation on upstream DNA can cause RNAP forward motion, Mfd needs to bind to the elongation complex. The most likely mode of recruitment of Mfd to paused or stalled elongation complexes is direct recruitment through interaction of the contact patches on Mfd and the β subunit. However, in theory the interaction between Mfd and RNAP could be formed at the promoter or during early elongation, resulting in Mfd being carried along with the elongation complex during normal transcription and being present to fulfil its role as a TRCF upon stalling of transcription. Alternatively, Mfd could be constantly translocating along the DNA searching for stalled elongation complexes. When the Mfd protein encounters a stalled elongation complex it could then interact with the RNAP and displace it from the DNA. Mfd was still able to displace an elongation complex with a protein roadblock on the upstream DNA, providing a 26 bp gap was present between the front edge of the roadblock and the rear edge of the elongation complex (Park *et al.*, 2002). Since the roadblock would have inhibited translocation along the DNA this argues against Mfd travelling along the DNA in order to find stalled elongation complexes.

Although the interaction between Mfd and RNAP core in pull-down and yeast two-hybrid experiments appears to be weak, the conformation of the β1 domain in an elongation complex could increase the affinity of the interaction with Mfd. In addition, Mfd binding to upstream DNA could help to stabilise the interaction with the elongation complex. The binding patch for Mfd on the β1 domain is located on the same face as the predicted exit point of upstream DNA from the elongation complex. As such one can envisage how an interaction between the Mfd RID and this region of the β subunit could bring the helicase domains of Mfd close to the upstream DNA. The data are consistent with there being a single interaction patch

between the RID of Mfd and the first 142 amino acids of the β subunit of RNAP. However, MBP-Mfd bound to beads was able to bind both wild-type RNAP and RNAP containing substitutions in the β subunit that abolished Mfd-mediated elongation complex displacement (Dr. A. Smith, personal communication). This observation could be an experimental artefact; although the affinity of Mfd for the mutant RNAP has been greatly reduced, the high concentration of RNAP used in the pull-down assays could remain above the K_d. Alternatively the data from the pull-downs (which use the full-length proteins) could indicate that there is a second interaction site in addition to the contact formed between the RID of Mfd and amino acids 1-142 of the β subunit.

Based on studies of Mfd and other helicase proteins, including the identification in this thesis of substitutions in three regions of Mfd that abolish displacement of RNAP, the following mechanistic model can be proposed to describe Mfd-mediated termination of transcription. Following recruitment of Mfd to a stalled elongation complex through an interaction between the RID of Mfd and the N-terminus of the β subunit, Mfd is able to bind and translocate on dsDNA upstream of the elongation complex. ATP binding and hydrolysis is coupled, through the TRG motif and changes in the relative orientations of the helicase domains, to DNA binding and release by a single DNA contact site in order to track along the backbone of the DNA helix. The result is that RNA polymerase is pushed forwards and if transcription is unable to continue, this results in shortening and destabilisation of the hybrid. The destabilised elongation complex will be removed from the DNA template, presumably requiring opening of the jaws of RNAP, a process that may also be assisted by Mfd.

The mechanism of transcription-coupled repair appears to be different in eukaryotes than in prokaryotes. CSB is not able to disrupt ternary transcription complexes but is a SF2 helicase, which is stimulated by DNA and is able to bind to components of the transcriptional machinery (Selby and Sancar, 1997b). CSB may be able to move stalled elongation complexes on the DNA template to allow access of repair proteins rather than completely removing RNA polymerase. It also appears that a second pathway is present in eukaryotes to remove stalled elongation complexes if action by CSB fails. In yeast the Def1 protein forms a complex with the CSB homolog, Rad26. When a lesion cannot be quickly removed by Rad26-stimulated repair, Def1 has been shown to be capable of causing ubiquitination of RNA polymerase II, targeting it for degradation by the proteasome (Woudstra *et al.*, 2002).

Further questions

Do Mfd and σ region 3 compete for binding to the β1 domain of RNAP?

A region of the β subunit, close to residues that interact with Mfd, also binds to σ region 3 (Murakami and Darst, 2003). In the crystal structure of the holoenzyme from *T. aquaticus* region σ_3 of σ^{70} is within 4 Å of amino acids 113-117 of the β subunit (Murakami *et al.*, 2002). These residues are equivalent to amino acids 123-127 of the β subunit of *E. coli* RNAP and despite not overlapping with the Mfd contact site; they are only 4 residues away from the Mfd contact patch. In the *T. thermophilus* holoenzyme crystal structure the terminal carbon atom of the E281 side chain of σ is only 5.9 Å away from the terminal carbon of the E110 side chain of the β1 Mfd contact patch (Vassylyev *et al.*, 2002). Therefore given the proximity of the Mfd contact site with that of σ_3 one could imagine that σ^{70} bound to the core

enzyme might preclude Mfd binding, explaining the inability of Mfd to displace σ-containing initiation complexes.

However, in pull-down experiments, complexes containing σ were able to bind to Mfd. It could be that in these complexes the contact between σ_3 and this region of the β subunit had been lost but that σ remained bound to other surfaces of the core enzyme. Therefore the reason why initiation complexes are Mfd-resistant could be because σ region 4.2 contacts the upstream DNA, which is also required by Mfd for RNAP displacement (Park *et al.*, 2002). Since Mfd is able to bind to transcription complexes containing σ, recruitment of Mfd to RNAP at the promoter cannot be dismissed.

<u>What is the fate of stalled transcription initiation complexes?</u>

Mfd is only able to promote transcription-coupled repair of DNA at least 15 bp downstream of the promoter (Selby and Sancar, 1995b), presumably due to its inability to displace initiation complexes from the DNA (Selby and Sancar, 1995a). Therefore how is damage close to the promoter repaired? Since initiation complexes have been shown by footprinting to cover up to approximately +20 (Li, 1998), it seems probable that lesions located upstream of +15 prevent binding of RNAP to the promoter, although this has not been examined to date. This would mean that lesions near the promoter would not be protected by bound RNAP and therefore would be accessible to repair enzymes for global NER. Initiation complexes are more salt-sensitive and therefore less stable than elongation complexes. If a lesion did not prevent RNAP binding to a promoter, it could cause spontaneous dissociation of the initiation complex when it was encountered.

What is the fate of the Mfd protein following RNAP removal?

Following displacement of an elongation complex, it is not known whether the Mfd protein also leaves the DNA or whether it replaces RNAP on the DNA template. One might assume that in order to recruit UvrA to the lesion, Mfd remains bound to the DNA after it has brought about RNAP dissociation and therefore the process of RNAP displacement would also involve breaking the interaction between Mfd and RNAP, although no such Mfd-DNA intermediate has yet been detected (Selby and Sancar, 1993b). However if UvrA has been recruited to the site prior to RNAP displacement there would be no requirement for Mfd to stay bound to the DNA. In addition there is no reason why Mfd needs to remain associated with the DNA following rescue of backtracked elongation complexes. It may be possible to address the question of whether Mfd remains bound to the DNA following elongation complex displacement using rapid footprinting techniques.

DNA binding studies suggested that the Mfd protein is capable of wrapping DNA ((Selby and Sancar, 1995a) and chapter 5 of this thesis). Only 26 bp of upstream DNA is required for elongation complex displacement, too little for extensive DNA wrapping and as such it seems unlikely that wrapping is necessary for the RNA displacement activity of the Mfd protein. One possible role for wrapping of DNA by Mfd is during a later stage of the transcription-coupled repair pathway, such as remaining bound to DNA following displacement or in loading of UvrA onto the DNA. It has been shown that 90 bp of DNA is needed downstream of a lesion in order for it to be subject to transcription-coupled repair (Selby and Sancar, 1995b) and this would be consistent with a role in the process for DNA wrapping. It has

recently been shown by scanning force microscopy that the human CSB protein also wraps DNA in a manner that requires ATP binding (Beerens *et al.*, 2004).

How does the Mfd protein recruit UvrA?

The mechanism of UvrA recruitment during transcription-coupled repair has not been addressed in detail in the published literature. It is assumed that Mfd interacts with UvrA through a section in the N-terminus that shows homology to UvrB. The residues critical for this interaction are not known and it is unknown whether Mfd bringing UvrA close to the lesion site is sufficient to result in strand specific repair or if the UvrA is actually loaded onto the non-transcribed strand by Mfd and then moves 5′-3′ to search for damage on the transcribed strand. One model for transcription-coupled repair proposes that $UvrA_2B$ is loaded onto genes that are being transcribed through interaction of UvrA with RNAP, recruiting $UvrA_2B$ onto open promoters (Ahn and Grossman, 1996). The Uvr helicase would be loaded onto the non-transcribed strand to search for damage in the template strand ahead of the transcribing RNAP. This pathway negates the need for Mfd. A mechanistic model incorporating both Mfd-dependent and $UvrA_2B$ promoter binding pathways of transcription-coupled repair has been constructed (Patel *et al.*, 2004). This model indicates that when a high level of recruitment to the promoter is possible, repair will occur in an Mfd-independent manner, while Mfd-dependent repair will occur if binding of $UvrA_2B$ to the promoter is impeded. However there are only 10 copies of UvrA in a cell, rising to 150 copies upon induction of the SOS response (Grossman and Thiagalingam, 1993), compared to 500 copies of Mfd (Selby and Sancar, 1993a) and there is a clear defect in transcription-coupled repair in *mfd*⁻ cells.

MfdTrunc$_{1\text{-}997}$, which lacks the C-terminal 151 amino acids of the wild-type protein, was able to displace elongation complexes *in vivo* and has been shown to also be capable of displacing elongation complexes *in vitro*. This protein increased the UV sensitivity of an *mfd*⁻ strain (Dr. A. Smith, personal communication), a phenomenon which had previously been observed using MfdTrunc$_{1\text{-}939}$ (Selby and Sancar, 1995a). It was proposed that the increase in UV sensitivity upon introduction of MfdTrunc$_{1\text{-}939}$ was due to sequestration of UvrA. MfdTrunc$_{1\text{-}939}$ was also shown in EMSAs to interfere with UvrA binding to damaged DNA. The region that has been deleted in MfdTrunc$_{1\text{-}997}$ may mask the UvrA interaction site of Mfd in the wild-type protein until a conformational change is induced either by RNAP binding or RNAP displacement. Studying the effect of MfdTrunc$_{1\text{-}997}$ on DNA binding by UvrA and in nucleotide excision repair assays in the presence or absence of an excess of the C-terminal 151 amino acid fragment may help to examine this model. The truncated Mfd protein could be used to study the interaction of Mfd with UvrA and the competition/handover between Mfd and UvrB for binding to UvrA.

<u>What are the links between Mfd and the BER repair pathway?</u>

Transcription-coupled repair has been proposed to act on oxidative lesions, in addition to its well-documented action on NER substrates, and this process may involve the Mfd protein (Bregeon *et al.*, 2003). In a strain with a null mutation in the *mfd* gene, transcriptional bypass of an oxidative lesion led to the generation of mutated transcripts. This was also observed in a strain containing a null mutation in the *mutM* gene (encoding the enzyme responsible for removal of 8-oxoG). The effect was greater in a strain that contained both of these null mutations. The authors interpret these data as indicating that 8-oxoG is subject to transcription-coupled

repair and that BER and TCR may compete to repair lesions on the template strand, however the assay that they use is an assay of reporter activity rather than an assay for repair *per se*. Therefore it could be that the effect on transcription of a reporter gene downstream of an 8-oxoG lesion seen in *mfd* cells is due to a lack of Mfd-mediated displacement of RNAP. Whereas pyrimidine dimers cause stalling of elongation complexes to act as the signal for TCR, this cannot be the case for TCR of oxidative lesions. Slowing of the elongation complex by the lesion (or intermediates in the repair of the lesion) may be sufficient to act as a signal for preferential repair or Mfd-mediated displacement of RNAP. It would be of interest to study more fully the effects of different lesions on the rate of transcription and whether these rates correlate with the ability of Mfd to displace RNAP at these sites. As the effect of the lesion on elongation increases, one would expect the balance between Mfd-mediated termination and continued elongation to be tipped in favour of termination.

REFERENCES

Abo, T., Inada, T., Ogawa, K. and Aiba, H. (2000) SsrA-mediated tagging and proteolysis of LacI and its role in the regulation of *lac* operon. *EMBO J*, **19**, 3762-3769.

Adelman, K., Yuzenkova, J., La Porta, A., Zenkin, N., Lee, J., Lis, J.T., Borukhov, S., Wang, M.D. and Severinov, K. (2004) Molecular mechanism of transcription inhibition by peptide antibiotic microcin j25. *Mol Cell*, **14**, 753-762.

Ahn, B. and Grossman, L. (1996) RNA polymerase signals UvrAB landing sites. *J Biol Chem*, **271**, 21453-21461.

Auble, D.T. and Steggerda, S.M. (1999) Testing for DNA tracking by MOT1, a SNF2/SWI2 protein family member. *Mol Cell Biol*, **19**, 412-423.

Auble, D.T., Wang, D., Post, K.W. and Hahn, S. (1997) Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. *Mol Cell Biol*, **17**, 4842-4851.

Ayora, S., Rojo, F., Ogasawara, N., Nakai, S. and Alonso, J.C. (1996) The Mfd protein of *Bacillus subtilis* 168 is involved in both transcription-coupled DNA repair and DNA recombination. *J Mol Biol*, **256**, 301-318.

Bahr, A., De Graeve, F., Kedinger, C. and Chatton, B. (1998) Point mutations causing Bloom's syndrome abolish ATPase and DNA helicase activities of the BLM protein. *Oncogene*, **17**, 2565-2571.

Bar-Nahum, G. and Nudler, E. (2001) Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *Escherichia coli*. *Cell*, **106**, 443-451.

Barne, K.A., Bown, J.A., Busby, S.J. and Minchin, S.D. (1997) Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J*, **16**, 4034-4040.

Beerens, N., Hoeijmakers, J.H., Kanaar, R., Vermeulen, W. and Wyman, C. (2004) The CSB protein actively wraps DNA. *J Biol Chem*. **280**, 4722-4729

Bell, C.E. and Lewis, M. (2001) Crystallographic analysis of Lac repressor bound to natural operator O1. *J Mol Biol*, **312**, 921-926.

Bernstein, D.A., Zittel, M.C. and Keck, J.L. (2003) High-resolution structure of the *E.coli* RecQ helicase catalytic core. *EMBO J*, **22**, 4910-4921.

Bockrath, R.C. and Palmer, J.E. (1977) Differential repair of premutational UV-lesions at tRNA genes in *E. coli*. *Mol Gen Genet*, **156**, 133-140. **280**, 4722-4729

Bogden, C.E., Fass, D., Bergman, N., Nichols, M.D. and Berger, J.M. (1999) The structural basis for terminator recognition by the Rho transcription termination factor. *Mol Cell*, **3**, 487-493.

Bohr, V.A., Smith, C.A., Okumoto, D.S. and Hanawalt, P.C. (1985) DNA repair in an active gene: removal of pyrimidine dimers from the *DHFR* gene of CHO cells is much more efficient than in the genome overall. *Cell*, **40**, 359-369.

Bondarenko, V.A., Jiang, Y.I. and Studitsky, V.M. (2003) Rationally designed insulator-like elements can block enhancer action in vitro. *EMBO J*, **22**, 4728-4737.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248-254.

Bregeon, D., Doddridge, Z.A., You, H.J., Weiss, B. and Doetsch, P.W. (2003) Transcriptional mutagenesis induced by uracil and 8-oxoguanine in *Escherichia coli*. *Mol Cell*, **12**, 959-970.

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D. and Heumann, H. (2004) The σ^{70} subunit of RNA polymerase induces *lacUV5* promoter-proximal pausing of transcription. *Nat Struct Mol Biol*, **11**, 551-557.

Burns, H. and Minchin, S. (1994) Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. *Nucleic Acids Res*, **22**, 3840-3845.

Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L. and Gottesman, M.E. (1995) *Escherichia coli* NusG protein stimulates transcription elongation rates *in vivo* and *in vitro*. *J Bacteriol*, **177**, 1388-1392.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. and Darst, S.A. (2001) Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell*, **104**, 901-912.

Campbell, F.E., Jr. and Setzer, D.R. (1991) Displacement of *Xenopus* transcription factor IIIA from a 5S rRNA gene by a transcribing RNA polymerase. *Mol Cell Biol*, **11**, 3978-3986.

Carell, T., Burgdorf, L.T., Kundu, L.M. and Cichon, M. (2001) The mechanism of action of DNA photolyases. *Curr Opin Chem Biol*, **5**, 491-498.

Caron, P.R. and Grossman, L. (1988) Involvement of a cryptic ATPase activity of UvrB and its proteolysis product, UvrB* in DNA repair. *Nucleic Acids Res*, **16**, 10891-10902.

Caruthers, J.M., Johnson, E.R. and McKay, D.B. (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A*, **97**, 13080-13085.

Caruthers, J.M. and McKay, D.B. (2002) Helicase structure and mechanism. *Curr Opin Struct Biol*, **12**, 123-133.

Chambers, A.L., Smith, A.J. and Savery, N.J. (2003) A DNA translocation motif in the bacterial transcription-repair coupling factor, Mfd. *Nucleic Acids Res*, **31**, 6409-6418.

Christiansen, M., Stevnsner, T., Modin, C., Martensen, P.M., Brosh, R.M., Jr and Bohr, V.A. (2003) Functional consequences of mutations in the conserved SF2 motifs and post-translational phosphorylation of the CSB protein. *Nucl. Acids. Res.*, **31**, 963-973.

Claassen, L.A. and Grossman, L. (1991) Deletion mutagenesis of the *Escherichia coli* UvrA protein localizes domains for DNA binding, damage recognition, and protein-protein interactions. *J Biol Chem*, **266**, 11388-11394.

Cline, J., Braman, J.C. and Hogrefe, H.H. (1996) PCR fidelity of *pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res*, **24**, 3546-3551.

Cordin, O., Tanner, N.K., Doere, M., Linder, P. and Banroques, J. (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J*, **23**, 2478-2487.

Darst, R.P., Wang, D. and Auble, D.T. (2001) MOT1-catalyzed TBP-DNA disruption: uncoupling DNA conformational change and role of upstream DNA. *EMBO J.*, **20**, 2028-2040.

Darst, S.A. (2001) Bacterial RNA polymerase. *Curr Opin Struct Biol*, **11**, 155-162.

deHaseth, P.L., Zupancic, M.L. and Record, M.T., Jr. (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol*, **180**, 3019-3025.

Deng, L. and Shuman, S. (1998) Vaccinia NPH-I, a DExH-box ATPase, is the energy coupling factor for mRNA transcription termination. *Genes Dev*, **12**, 538-546.

Deuschle, U., Gentz, R. and Bujard, H. (1986) *lac* Repressor blocks transcribing RNA polymerase and terminates transcription. *Proc Natl Acad Sci U S A*, **83**, 4134-4137.

Dewitt, S.K. and Adelberg, E.A. (1962) Transduction of the attached sex factor of *Escherichia coli*. *J Bacteriol*, **83**, 673-678.

Dizdaroglu, M. (2003) Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. *Mutat Res*, **531**, 109-126.

Dole, S., Nagarajavel, V. and Schnetz, K. (2004) The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli bgl* operon downstream of the promoter. *Mol Microbiol*, **52**, 589-600.

Epshtein, V. and Nudler, E. (2003) Cooperation between RNA polymerase molecules in transcription elongation. *Science*, **300**, 801-805.

Epshtein, V., Toulme, F., Rahmouni, A.R., Borukhov, S. and Nudler, E. (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J*, **22**, 4719-4727.

Firman, K. and Szczelkun, M.D. (2000) Measuring motion on DNA by the type I restriction endonuclease *Eco*R124I using triplex displacement. *EMBO J*, **19**, 2094-2102.

Fitzgerald, D.J., DeLuca, C., Berger, I., Gaillard, H., Sigrist, R., Schimmele, K. and Richmond, T.J. (2004) Reaction cycle of the yeast Isw2 chromatin remodeling complex. *EMBO J*, **23**, 3836-3843.

Flick, J.S. and Johnston, M. (1990) Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae. *Mol Cell Biol*, **10**, 4757-4769.

Foster, J.E., Holmes, S.F. and Erie, D.A. (2001) Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell*, **106**, 243-252.

Friedberg, E.C., Walker, G.C. and Siede, W. (1995) Nucleotide excision repair in prokaryotes. In *DNA repair and mutagenesis*. American Society for Microbiology, Washington, DC 20005, pp. 191-232.

Friedman, A.M., Fischmann, T.O. and Steitz, T.A. (1995) Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science*, **268**, 1721-1727.

Gallivan, J.P. and McGarvey, M.J. (2003) The importance of the Q motif in the ATPase activity of a viral helicase. *FEBS Lett*, **554**, 485-488.

Geiselmann, J., Seifried, S.E., Yager, T.D., Liang, C. and von Hippel, P.H. (1992a) Physical properties of the *Escherichia coli* transcription termination factor rho. 2. Quaternary structure of the rho hexamer. *Biochemistry*, **31**, 121-132.

Geiselmann, J., Yager, T.D., Gill, S.C., Calmettes, P. and von Hippel, P.H. (1992b) Physical properties of the *Escherichia coli* transcription termination factor rho. 1. Association states and geometry of the rho hexamer. *Biochemistry*, **31**, 111-121.

Gorbalenya, A.E. and Koonin, E.V. (1993) Helicases - amino-acid-sequence comparisons and structure- function-Relationships. *Current Opinion in Structural Biology*, **3**, 419-429.

Gotta, S.L., Miller, O.L., Jr. and French, S.L. (1991) rRNA transcription rate in Escherichia coli. *J Bacteriol*, **173**, 6647-6649.

Greene J, M. (2002) Random mutagenesis by PCR. In F.M.Ausbel (ed.), *Short protocols in molecular biology: a compendium for Current protocols in moleucular biology*. John Wiley and Sons, Inc., pp. 8-7-8-12.

Grossman, L. and Thiagalingam, S. (1993) Nucleotide excision repair, a tracking mechanism in search of damage. *J Biol Chem*, **268**, 16871-16874.

Gruber, T.M., Markov, D., Sharp, M.M., Young, B.A., Lu, C.Z., Zhong, H.J., Artsimovitch, I., Geszvain, K.M., Arthur, T.M., Burgess, R.R., Landick, R., Severinov, K. and Gross, C.A. (2001) Binding of the initiation factor σ^{70} to core RNA polymerase is a multistep process. *Mol Cell*, **8**, 21-31.

Gusarov, I. and Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol Cell*, **3**, 495-504.

Gusarov, I. and Nudler, E. (2001) Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell*, **107**, 437-449.

Hall, M.C. and Matson, S.W. (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol*, **34**, 867-877.

Hara, R., Selby, C.P., Liu, M., Price, D.H. and Sancar, A. (1999) Human transcription release factor 2 dissociates RNA polymerases I and II stalled at a cyclobutane thymine dimer. *J Biol Chem*, **274**, 24779-24786.

He, B. and Zalkin, H. (1992) Repression of *Escherichia coli purB* is by a transcriptional roadblock mechanism. *J Bacteriol*, **174**, 7121-7127.

Hsu, L.M., Vo, N.V., Kane, C.M. and Chamberlin, M.J. (2003) In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. *Biochemistry*, **42**, 3777-3786.

Inacio, J.M., Costa, C. and de Sa-Nogueira, I. (2003) Distinct molecular mechanisms involved in carbon catabolite repression of the arabinose regulon in *Bacillus subtilis*. *Microbiology*, **149**, 2345-2355.

Jokerst, R.S., Weeks, J.R., Zehring, W.A. and Greenleaf, A.L. (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila. *Mol Gen Genet*, **215**, 266-275.

Kalogeraki, V.S., Tornaletti, S. and Hanawalt, P.C. (2003) Transcription arrest at a lesion in the transcribed DNA strand *in vitro* is not affected by a nearby lesion in the opposite strand. *J Biol Chem*, **278**, 19558-19564.

Kaplan, D.L. and O'Donnell, M. (2003) Rho factor: transcription termination in four steps. *Curr Biol*, **13**, R714-716.

Karreman, C. (1998) Fusion PCR, a one-step variant of the "megaprimer" method of mutagenesis. *Biotechniques*, **24**, 736-742.

Kashlev, M., Nudler, E., Goldfarb, A., White, T. and Kutter, E. (1993) Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell*, **75**, 147-154.

Kawaoka, J.J., E. Pyle, A.M. (2004) Backbone tracking by the SF2 helicase NPH-II. *Nat Struct Mol Biol*, **11**, 526-530.

Kim, H.C. and Gottesman, M.E. (2004) Transcription termination by phage HK022 Nun is facilitated by COOH-terminal lysine residues. *J Biol Chem*, **279**, 13412-13417.

King, R.A., Markov, D., Sen, R., Severinov, K. and Weisberg, R.A. (2004) A conserved zinc binding domain in the largest subunit of DNA-dependent RNA polymerase modulates intrinsic transcription termination and antitermination but does not stabilize the elongation complex. *J Mol Biol*, **342**, 1143-1154.

Komissarova, N., Becker, J., Solter, S., Kireeva, M. and Kashlev, M. (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell*, **10**, 1151-1162.

Komissarova, N. and Kashlev, M. (1997) Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A*, **94**, 1755-1760.

Komissarova, N. and Kashlev, M. (1998) Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc Natl Acad Sci U S A*, **95**, 14699-14704.

Kramer, H., Amouyal, M., Nordheim, A. and Muller-Hill, B. (1988) DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor. *EMBO J*, **7**, 547-556.

Landick, R. (1997) RNA polymerase slides home: pause and termination site recognition. *Cell*, **88**, 741-744.

Laptenko, O., Lee, J., Lomakin, I. and Borukhov, S. (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J*, **22**, 6322-6334.

Levin, J.R., Krummel, B. and Chamberlin, M.J. (1987) Isolation and properties of transcribing ternary complexes of *Escherichia coli* RNA polymerase positioned at a single template base. *J Mol Biol*, **196**, 85-100.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G. and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, **271**, 1247-1254.

Li, J., Horwitz, R., McCracken, S. and Greenblatt, J. (1992) NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. *J Biol Chem*, **267**, 6012-6019.

Li, X.Y.M., W.R. (1998) Characterisation of the closed complex intermediate formed during transcription initiation by *Escherichia coli* RNA polymerase. *J Biol Chem*, **273**, 23549-23557.

Liu, J. and Doetsch, P.W. (1996) Template strand gap bypass is a general property of prokaryotic RNA polymerases: implications for elongation mechanisms. *Biochemistry*, **35**, 14999-15008.

Lloyd, R.G. and Sharples, G.J. (1993) Dissociation of synthetic Holliday junctions by *E. coli* RecG protein. *EMBO J*, **12**, 17-22.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N.R. (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett*, **74**, 271-276.

Machius, M., Henry, L., Palnitkar, M. and Deisenhofer, J. (1999) Crystal structure of the DNA nucleotide excision repair enzyme UvrB from *Thermus thermophilus*. *Proc Natl Acad Sci U S A*, **96**, 11717-11722.

Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q. and Lloyd, R.G. (2003) A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. *EMBO J.*, **22**, 724-734.

McGlynn, P. and Lloyd, R.G. (2002) Genome stability and the processing of damaged replication forks by RecG. *Trends Genet*, **18**, 413-419.

Mellon, I. and Hanawalt, P.C. (1989) Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature*, **342**, 95-98.

Mellon, I., Spivak, G. and Hanawalt, P.C. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian *DHFR* gene. *Cell*, **51**, 241-249.

Michalke, H. and Bremer, H. (1969) RNA synthesis in *Escherichia coli* after irradiation with ultraviolet light. *J Mol Biol*, **41**, 1-23.

Miller, J. (1972) In *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press.

Mukhopadhyay, J., Kapanidis, A.N., Mekler, V., Kortkhonjia, E., Ebright, Y.W. and Ebright, R.H. (2001) Translocation of σ^{70} with RNA polymerase during

transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, **106**, 453-463.

Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R.M. and Ebright, R.H. (2004) Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Mol Cell*, **14**, 739-751.

Muller, J., Oehler, S. and Muller-Hill, B. (1996) Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J Mol Biol*, **257**, 21-29.

Muller-Hill, B. (1998) The function of auxiliary operators. *Mol Microbiol*, **29**, 13-18.

Murakami, K.S. and Darst, S.A. (2003) Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology*, **13**, 31-39.

Murakami, K.S., Masuda, S. and Darst, S.A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, **296**, 1280-1284.

Nickels, B.E., Roberts, C.W., Sun, H., Roberts, J.W. and Hochschild, A. (2002) The σ^{70} subunit of RNA polymerase is contacted by the (lambda)Q antiterminator during early elongation. *Mol Cell*, **10**, 611-622.

Niu, W., Kim, Y., Tau, G., Heyduk, T. and Ebright, R.H. (1996) Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell*, **87**, 1123-1134.

Nudler, E., Avetissova, E., Markovtsov, V. and Goldfarb, A. (1996) Transcription processivity: protein-DNA interactions holding together the elongation complex. *Science*, **273**, 211-217.

Nudler, E. and Gottesman, M.E. (2002) Transcription termination and anti-termination in *E. coli*. *Genes Cells*, **7**, 755-768.

Nudler, E., Gusarov, I., Avetissova, E., Kozlov, M. and Goldfarb, A. (1998) Spatial organization of transcription elongation complex in *Escherichia coli*. *Science*, **281**, 424-428.

Nudler, E., Kashlev, M., Nikiforov, V. and Goldfarb, A. (1995) Coupling between transcription termination and RNA polymerase inchworming. *Cell*, **81**, 351-357.

Nudler, E., Mustaev, A., Lukhtanov, E. and Goldfarb, A. (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, **89**, 33-41.

Oehler, S., Eismann, E.R., Kramer, H. and Muller-Hill, B. (1990) The three operators of the *lac* operon cooperate in repression. *EMBO J*, **9**, 973-979.

Oh, E.Y., Claassen, L., Thiagalingam, S., Mazur, S. and Grossman, L. (1989) ATPase activity of the UvrA and UvrAB protein complexes of the Escherichia coli UvrABC endonuclease. *Nucleic Acids Res*, **17**, 4145-4159.

Oh, E.Y. and Grossman, L. (1987) Helicase properties of the Escherichia coli UvrAB protein complex. *Proc Natl Acad Sci U S A*, **84**, 3638-3642.

Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W. and Darst, S.A. (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell*, **114**, 335-345.

Padmanabhan, S., Elias-Arnanz, M., Carpio, E., Aparicio, P. and Murillo, F.J. (2001) Domain architecture of a high mobility group A-type bacterial transcriptional factor. *J Biol Chem*, **276**, 41566-41575.

Park, J.S., Marr, M.T. and Roberts, J.W. (2002) *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell*, **109**, 757-767.

Patel, S., Venkatesh, K.V. and Edwards, J.S. (2004) An integrated mechanistic model for transcription-coupled nucleotide excision repair. *DNA Repair (Amst)*, **3**, 343-348.

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W. and Gourse, R.L. (2004) DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell*, **118**, 311-322.

Pavco, P.A. and Steege, D.A. (1990) Elongation by *Escherichia coli* RNA polymerase is blocked *in vitro* by a site-specific DNA binding protein. *J Biol Chem*, **265**, 9960-9969.

Pavco, P.A. and Steege, D.A. (1991) Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein. *Nucleic Acids Res*, **19**, 4639-4646.

Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I. and Vassylyev, D.G. (2004) Regulation through the secondary channel-structural framework for ppGpp-DksA synergism during transcription. *Cell*, **118**, 297-309.

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A. and Legrain, P. (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature*, **409**, 211-215.

Reece, R.J. and Maxwell, A. (1991) The C-terminal domain of the *Escherichia coli* DNA gyrase A subunit is a DNA-binding protein. *Nucleic Acids Res*, **19**, 1399-1405.

Reines, D. and Mote, J., Jr. (1993) Elongation factor SII-dependent transcription by RNA polymerase II through a sequence-specific DNA-binding protein. *Proc Natl Acad Sci U S A*, **90**, 1917-1921.

Richardson, J.P. (1996) Structural organization of transcription termination factor Rho. *J Biol Chem*, **271**, 1251-1254.

Richmond, E. and Peterson, C.L. (1996) Functional analysis of the DNA-stimulated ATPase domain of yeast SWI2/SNF2. *Nucleic Acids Res*, **24**, 3685-3692.

Ryder, A.M. and Roberts, J.W. (2003) Role of the non-template strand of the elongation bubble in intrinsic transcription termination. *J Mol Biol*, **334**, 205-213.

Sambrook, J. and Russel, D.W. (2001) *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Santangelo, T.J. and Roberts, J.W. (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol Cell*, **14**, 117-126.

Savery, N.J., Lloyd, G.S., Busby, S.J., Thomas, M.S., Ebright, R.H. and Gourse, R.L. (2002) Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase alpha subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters. *J Bacteriol*, **184**, 2273-2280.

Seeley, T.W. and Grossman, L. (1990) The role of Escherichia coli UvrB in nucleotide excision repair. *J Biol Chem*, **265**, 7158-7165.

Selby, C.P., Drapkin, R., Reinberg, D. and Sancar, A. (1997) RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic Acids Res*, **25**, 787-793.

Selby, C.P. and Sancar, A. (1990) Transcription preferentially inhibits nucleotide excision repair of the template DNA strand *in vitro*. *J Biol Chem*, **265**, 21330-21336.

Selby, C.P. and Sancar, A. (1991) Gene- and strand-specific repair in vitro: partial purification of a transcription-repair coupling factor. *Proc Natl Acad Sci U S A*, **88**, 8232-8236.

Selby, C.P. and Sancar, A. (1993a) Molecular mechanism of transcription-repair coupling. *Science*, **260**, 53-58.

Selby, C.P. and Sancar, A. (1993b) Transcription-repair coupling and mutation frequency decline. *J Bacteriol*, **175**, 7509-7514.

Selby, C.P. and Sancar, A. (1995a) Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *J Biol Chem*, **270**, 4882-4889.

Selby, C.P. and Sancar, A. (1995b) Structure and function of transcription-repair coupling factor. II. Catalytic properties. *J Biol Chem*, **270**, 4890-4895.

Selby, C.P. and Sancar, A. (1997a) Cockayne syndrome group B protein enhances elongation by RNA polymerase II. *Proc Natl Acad Sci U S A*, **94**, 11205-11209.

Selby, C.P. and Sancar, A. (1997b) Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *J Biol Chem*, **272**, 1885-1890.

Selby, C.P., Witkin, E.M. and Sancar, A. (1991) *Escherichia coli mfd* mutant deficient in "mutation frequency decline" lacks strand-specific repair: *in vitro* complementation with purified coupling factor. *Proc Natl Acad Sci U S A*, **88**, 11574-11578.

Sellitti, M.A., Pavco, P.A. and Steege, D.A. (1987) *lac* repressor blocks *in vivo* transcription of *lac* control region DNA. *Proc Natl Acad Sci U S A*, **84**, 3199-3203.

Severinov, K., Kashlev, M., Severinova, E., Bass, I., McWilliams, K., Kutter, E., Nikiforov, V., Snyder, L. and Goldfarb, A. (1994) A non-essential domain of *Escherichia coli* RNA polymerase required for the action of the termination factor Alc. *J Biol Chem*, **269**, 14254-14259.

Shi, Y.B., Gamper, H. and Hearst, J.E. (1987) The effects of covalent additions of a psoralen on transcription by *E. coli* RNA polymerase. *Nucleic Acids Res*, **15**, 6843-6854.

Shi, Y.B., Gamper, H. and Hearst, J.E. (1988a) Interaction of T7 RNA polymerase with DNA in an elongation complex arrested at a specific psoralen adduct site. *J Biol Chem*, **263**, 527-534.

Shi, Y.B., Gamper, H., Van Houten, B. and Hearst, J.E. (1988b) Interaction of *Escherichia coli* RNA polymerase with DNA in an elongation complex arrested at a specific psoralen crosslink site. *J Mol Biol*, **199**, 277-293.

Shin, D.H., Nguyen, H.H., Jancarik, J., Yokota, H., Kim, R. and Kim, S.H. (2003) Crystal structure of NusA from *Thermotoga maritima* and functional implication of the N-terminal domain. *Biochemistry*, **42**, 13429-13437.

Simons, A., Tils, D., von Wilcken-Bergmann, B. and Muller-Hill, B. (1984) Possible ideal *lac* operator: *Escherichia coli lac* operator-like sequences from eukaryotic genomes lack the central G:C pair. *Proc Natl Acad Sci U S A*, **81**, 1624-1628.

Singleton, M.R., Scaife, S. and Wigley, D.B. (2001) Structural analysis of DNA replication fork reversal by RecG. *Cell*, **107**, 79-89.

Singleton, M.R. and Wigley, D.B. (2003) Multiple roles for ATP hydrolysis in nucleic acid modifying enzymes. *EMBO J*, **22**, 4579-4583.

Slijper, M., Bonvin, A.M., Boelens, R. and Kaptein, R. (1996) Refined structure of lac repressor headpiece (1-56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the lac operator. *J Mol Biol*, **259**, 761-773.

Snyder, L., Gold, L. and Kutter, E. (1976) A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. *Proc Natl Acad Sci U S A*, **73**, 3098-3102.

Stanley, L.K. and Savery, N.J. (2003) Characterisation of the *Escherichia coli mfd* promoter. *Arch Microbiol*, **179**, 381-385.

Stitt, B.L. (1988) *Escherichia coli* transcription termination protein rho has three hydrolytic sites for ATP. *J Biol Chem*, **263**, 11130-11137.

Subramanya, H.S., Bird, L.E., Brannigan, J.A. and Wigley, D.B. (1996) Crystal structure of a DExx box DNA helicase. *Nature*, **384**, 379-383.

Sullivan, S.L. and Gottesman, M.E. (1992) Requirement for *E. coli* NusG protein in factor-dependent transcription termination. *Cell*, **68**, 989-994.

Svetlov, V., Vassylyev, D.G. and Artsimovitch, I. (2004) Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J Biol Chem*, **279**, 38087-38090.

Sweder, K.S. and Hanawalt, P.C. (1992) Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proc Natl Acad Sci U S A*, **89**, 10696-10700.

Sweetser, D., Nonet, M. and Young, R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A*, **84**, 1192-1196.

Tai, C.L., Pan, W.C., Liaw, S.H., Yang, U.C., Hwang, L.H. and Chen, D.S. (2001) Structure-based mutational analysis of the hepatitis C virus NS3 helicase. *J Virol*, **75**, 8289-8297.

Tanner, N.K., Cordin, O., Banroques, J., Doere, M. and Linder, P. (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell*, **11**, 127-138.

Theis, K., Chen, P.J., Skorvaga, M., Van Houten, B. and Kisker, C. (1999) Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J*, **18**, 6899-6907.

Theis, K., Skorvaga, M., Machius, M., Nakagawa, N., Van Houten, B. and Kisker, C. (2000) The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutat Res*, **460**, 277-300.

Thiagalingam, S. and Grossman, L. (1993) The multiple roles for ATP in the *Escherichia coli* UvrABC endonuclease-catalyzed incision reaction. *J Biol Chem*, **268**, 18382-18389.

Tombline, G. and Fishel, R. (2002) Biochemical characterization of the human RAD51 protein. I. ATP hydrolysis. *J Biol Chem*, **277**, 14417-14425.

Tornaletti, S. and Hanawalt, P.C. (1999) Effect of DNA lesions on transcription elongation. *Biochimie*, **81**, 139-146.

Tornaletti, S., Maeda, L.S., Kolodner, R.D. and Hanawalt, P.C. (2004) Effect of 8-oxoguanine on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. *DNA Repair (Amst)*, **3**, 483-494.

Tornaletti, S., Reines, D. and Hanawalt, P.C. (1999) Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. *J Biol Chem*, **274**, 24124-24130.

Toulokhonov, I., Artsimovitch, I. and Landick, R. (2001) Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science*, **292**, 730-733.

Van Houten, B. (1990) Nucleotide excision repair in Escherichia coli. *Microbiol Rev*, **54**, 18-51.

Van Houten, B., Gamper, H., Sancar, A. and Hearst, J.E. (1987) DNase I footprint of ABC excinuclease. *J Biol Chem*, **262**, 13180-13187.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*, **417**, 712-719.

Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S. and Wigley, D.B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, **97**, 75-84.

Verhoeven, E.E.A., Wyman, C., Moolenaar, G.F. and Goosen, N. (2002) The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *EMBO J.*, **21**, 4196-4205.

Virnik, K., Lyubchenko, Y.L., Karymov, M.A., Dahlgren, P., Tolstorukov, M.Y., Semsey, S., Zhurkin, V.B. and Adhya, S. (2003) "Antiparallel" DNA loop in gal repressosome visualized by atomic force microscopy. *J Mol Biol*, **334**, 53-63.

Viswanathan, A. and Doetsch, P.W. (1998) Effects of nonbulky DNA base damages on *Escherichia coli* RNA polymerase-mediated elongation and promoter clearance. *J Biol Chem*, **273**, 21276-21281.

Vogel, U. and Jensen, K.F. (1994) The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J Bacteriol*, **176**, 2807-2813.

von Hippel, P.H. (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science*, **281**, 660-665.

von Hippel, P.H., Bear, D.G., Morgan, W.D. and McSwiggen, J.A. (1984) Protein-nucleic acid interactions in transcription: a molecular analysis. *Annu Rev Biochem*, **53**, 389-446.

Washburn, R.S., Wang, Y. and Gottesman, M.E. (2003) Role of *E. coli* transcription-repair coupling factor Mfd in Nun-mediated transcription termination. *J Mol Biol*, **329**, 655-662.

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M.D. and Owen-Hughes, T. (2003) Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Mol Cell Biol*, **23**, 1935-1945.

Wilson, K.S., Conant, C.R. and von Hippel, P.H. (1999) Determinants of the stability of transcription elongation complexes: interactions of the nascent RNA with the DNA template and the RNA polymerase. *J Mol Biol*, **289**, 1179-1194.

Witkin, E.M. (1966) Radiation-induced mutations and their repair. *Science*, **152**, 1345-1353.

Woudstra, E.C., Gilbert, C., Fellows, J., Jansen, L., Brouwer, J., Erdjument-Bromage, H., Tempst, P. and Svejstrup, J.Q. (2002) A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature*, **415**, 929-933.

Wu, H.M. and Crothers, D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509-513.

Xie, Z. and Price, D. (1997) Drosophila factor 2, an RNA polymerase II transcript release factor, has DNA-dependent ATPase activity. *J Biol Chem*, **272**, 31902-31907.

Yang, W. (2000) Structure and function of mismatch repair proteins. *Mutat Res*, **460**, 245-256.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-119.

Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A.D., Le, H.V. and Weber, P.C. (1997) Structure of the hepatitis C virus RNA helicase domain. *Nat Struct Biol*, **4**, 463-467.

Yarnell, W.S. and Roberts, J.W. (1999) Mechanism of intrinsic transcription termination and antitermination. *Science*, **284**, 611-615.

Yin, H., Artsimovitch, I., Landick, R. and Gelles, J. (1999) Nonequilibrium mechanism of transcription termination from observations of single RNA polymerase molecules. *Proc Natl Acad Sci U S A*, **96**, 13124-13129.

Zalieckas, J.M., Wray, L.V., Jr., Ferson, A.E. and Fisher, S.H. (1998) Transcription-repair coupling factor is involved in carbon catabolite repression of the *Bacillus subtilis hut* and *gnt* operons. *Mol Microbiol*, **27**, 1031-1038.

Zaychikov, E., Denissova, L. and Heumann, H. (1995) Translocation of the Escherichia coli transcription complex observed in the registers 11 to 20: "jumping" of RNA polymerase and asymmetric expansion and contraction of the "transcription bubble". *Proc Natl Acad Sci U S A*, **92**, 1739-1743.

Zellars, M. and Squires, C.L. (1999) Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol Microbiol*, **32**, 1296-1304.

Zeng, X., Galinier, A. and Saxild, H.H. (2000) Catabolite repression of *dra-nupC-pdp* operon expression in *Bacillus subtilis*. *Microbiology*, **146**, 2901-2908.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, **98**, 811-824.

Zhou, W. and Doetsch, P.W. (1993) Effects of abasic sites and DNA single-strand breaks on prokaryotic RNA polymerases. *Proc Natl Acad Sci U S A*, **90**, 6601-6605.

APPENDIX

Primer	Sequence 5′ → 3′	Description
MFD001	GGCAAGCTTTCAGAGAT CTTCCAGCGAGGC	Introduces stop codon and *Hin*dIII sites downstream of codon 939. Anneals to coding strand.
MFD002	TCTGGCAATAGCGACGG CAT	Anneals to codons 1008-1014. Anneals to non-coding strand.
MFD003	CAGTTCACCCGCGCCGG CAATCTCCAGATCGTG	Mutates arginine 953 codon to an alanine codon. Anneals to coding strand.
MFD005	AGCTGCACCAGTTACGC GGT	Anneals to codons 895-901. Anneals to coding strand.
MFD006	CACGATCTGGAGATTGC CGGCGCGGGTGAACTG	Mutates arginine 953 codon to an alanine codon. Anneals to non-coding strand.
MFD011	CTGGCGTCTCAGATCTT CCAGCGACGGC	Introduces a *Bsm*AI site adjacent to the downstream *Bgl*II site such that a *Bgl*II end is generated on digestion with *Bsm*AI. Anneals to coding strand.
MFD012	CCGATGCACAAAAAGCT CTTGAAGCAATTGCC	Mutates arginine 929 codon to an alanine codon. Anneals to non-coding strand.
MFD013	GGCAATTGCTTCAAGAG CTTTTTGTGCATCGG	Mutates arginine 929 codon to an alanine codon. Anneals to coding strand.
MFD024	CTGCTTGGCGAAGAAGC AAGCGGCTCAATGG	Mutates glutamine 963 to an alanine codon. Anneals to non-coding strand.
MFD025	TTCCATTGAGCCGCTTG CTTCTTCGCCAAGCAG	Mutates glutamine 963 to an alanine codon. Anneals to coding strand.
MFD027	GCATTAATGGCCTGCGC CGCATCCGGCGTGG	Mutates glutamnine 605 to an alanine codon. Anneals to coding strand.
MFD028	CGTTTGAAACCACGCCG GCGCAGGCGCAGGCC	Mutates aspartate 604 to an alanine codon. Anneals to non-coding strand.
MFD029	GGCCTGCGCCTGCGCCG GCGTGGTTTCAAACG	Mutates aspartate 604 to an alanine codon. Anneals to coding strand.
MFD030	ATTAGCCGTTACGCAGG TGG	Anneals to codons 529-535.Anneals to non-coding strand.
MFD031	CTTCATCGACAATCAGC AGG	Anneals to codons 725-732. Anneals to coding strand.

MFD032	GCATATTGGTCAACCGG TGGTCCATCTGG	Introduces *Age*I site by silent mutation of codon 481. Anneals to non-coding strand.
MFD033	CCAGATGGACCACCGGT TGACCAATATGC	Introduces *Age*I site by silent mutation of codon 481. Anneals to coding strand.
MFD034	TTCCAGAAACTGCCAGA CCT	Anneals to codons 355-361. Anneals to non-coding strand.
MFD035	CCACGCCGGATGCGGCG CAGGCCATTAATGC	Mutates glutamnine 605 to an alanine codon. Anneals to coding strand.
MFD039	CGATACACTGATCGCTA ACCTTGCGGAACTG	Mutates arginine 472 to an alanine codon. Anneals to non-coding strand.
MFD040	CAGTTCCGCAAGGTTAG CGATCAGTGTATCG	Mutates arginine 472 to an alanine codon. Anneals to coding strand.
MFD041	CGTAACCTTGCGGCTCT GCATATTGGTC	Mutates glutamate 476 to an alanine codon. Anneals to non-coding strand.
MFD042	GACCAATATGCAGAGCC GCAAGGTTACG	Mutates glutamate 476 to an alanine codon. Anneals to coding strand.
MFD043	CAGCCGGTGGTCGCTCT GGAGCACGGC	Mutates histidine 485 to an alanine codon. Anneals to non-coding strand.
MFD044	GCCGTGCTCCAGAGCGA CCACCGGCTG	Mutates histidine 485 to an alanine codon. Anneals to coding strand.
MFD045	GGTCCATCTGGAGGCTG GCGTCGGTCG	Mutates histidine 488 to an alanine codon. Anneals to non-coding strand.
MFD046	CGACCGACGCCAGCCTC CAGATGGACC	Mutates histidine 488 to an alanine codon. Anneals to coding strand.
MFD047	GCACGGCGTCGGTGCTT ATGCCGGAATGACC	Mutates arginine 492 to an alanine codon. Anneals to non-coding strand.
MFD048	GGTCATTCCGGCATAAG CACCGACGCCGTGC	Mutates arginine 492 to an alanine codon. Anneals to coding strand.
MFD049	GGAATGACCACGCTGGC TGCGGGTGGCATTACTG G	Mutates glutmate 500 to an alanine codon. Anneals to non-coding strand.
MFD050	CCAGTAATGCCACCCGC AGCCAGCGTGGTCATTC C	Mutates glutmate 500 to an alanine codon. Anneals to coding strand.
MFD051	GGCATTACTGGCGCTTA TCTGATGCTCACC	Mutates glutamate 507 to an alanine codon. Anneals to non-coding strand.
MFD052	GGTGAGCATCAGATAAG CGCCAGTAATGCC	Mutates glutamate 507 to an alanine codon. Anneals to coding strand.

MFD053	GCCAACGACGCCGCTCT GTATGTTCCG	Mutates lysine 518 to an alanine codon. Anneals to non-coding strand.
MFD054	CGGAACATACAGAGCG GCGTCGTTGGC	Mutates lysine 518 to an alanine codon. Anneals to coding strand.
MFD055	GCATCTGATTAGCGCTT ACGCAGGTGGC	Mutates arginine 531 to an alanine codon. Anneals to non-coding strand.
MFD056	GCCACCTGCGTAAGCGC TAATCAGATGC	Mutates arginine 531 to an alanine codon. Anneals to coding strand.
MFD057	GCCCCGCTGCATGCTCT TGGCGGCGATGC	Mutates lysine 544 to an alanine codon. Anneals to non-coding strand.
MFD058	GCATCGCCGCCAAGAGC ATGCAGCGGGGC	Mutates lysine 544 to an alanine codon. Anneals to coding strand.
MFD059	CGATGCGTGGTCAGCTG CGCGGCAGAAAGC	Mutates arginine 552 to an alanine codon. Anneals to non-coding strand.
MFD060	GCTTTCTGCCGCGCAGC TGACCACGCATCG	Mutates arginine 552 to an alanine codon. Anneals to coding strand.
MFD061	CGTGGTCACGCGCGGCT CAGAAAGCGGCG	Mutates arginine 554 to an alanine codon. Anneals to non-coding strand.
MFD062	CGCCGCTTTCTGAGCCG CGCGTGACCACG	Mutates arginine 554 to an alanine codon. Anneals to coding strand.
MFD063	CGCGCGCGGCAGGCTGC GGCGGAAAAAGTGC	Mutates lysine 556 to an alanine codon. Anneals to non-coding strand.
MFD064	GCACTTTTTCCGCCGCA GCCTGCCGCGCGCG	Mutates lysine 556 to an alanine codon. Anneals to coding strand.
MFD065	CGGAAAAAGTGCGTGCC GTGGCGGCGGAATTGC	Mutates aspartate 563 to an alanine codon. Anneals to non-coding strand.
MFD066	GCAATTCCGCCGCCACG GCACGCACTTTTTCCG	Mutates aspartate 563 to an alanine codon. Anneals to coding strand.
MFD067	CGTGATGTGGCGGCGGC TTTGCTGGATATCTACG	Mutates glutamate 567 to an alanine codon. Anneals to non-coding strand.
MFD068	CGTAGATATCCAGCAAA GCCGCCGCCACATCACG	Mutates glutamate 567 to an alanine codon. Anneals to coding strand.
Luc1	GCGAATTCCGCCATGGA AGACGCCAAAAAC	Introduces *Nco*I site at ATG of *luc* gene of pGL2 and an *Eco*RI site upstream. Anneals to non-coding strand.
Luc2	CGCGAAGCTTACAATTT GGACTTTCC	Introduces *Hin*dIII site at the end of *luc* gene of pGL2. Anneals to coding strand.

Luc3	GTAGATCCAGAGGAGTT CATTATCAGTG	Silently destroys *Eco*RI site within *luc* gene of pGL2. Anneals to coding strand.
Luc4	CACTGATAATGAACTCC TCTGGATCTAC	Silently destroys *Eco*RI site within *luc* gene of pGL2. Anneals to non-coding strand.
pSRSDNcoI	GTGATTTTTTTCTCCATG GTAGCTTCCTTAGCTC	Introduces *Nco*I site at ATG of *cat* gene. Contains Shine-Dalgarno sequence from pSRlacO$_{ID}$luc plasmid. Anneals to coding strand.
pSRds2	CACCGAAACGCCCAAG GCAGC	Introduces *Sty*I site downstream of *Bam*HI site in pSR*galPcon6*. Anneals to coding strand.
D5431	ACCTGACGTCTAAGAAA CC	Anneals 60 bp upstream of *Eco*RI site in pBR322-based plasmids.
Bluescript 1	TGACCATGATTACGCCA AGC	Anneals upstream of *Hind*III site of pBluescript KS+.
Bluescript 2	GCGCGTAATACGACTCA CTA	Anneals downstream of *Bam*HI site of BS+ KS.
M13 D9979 forward	GTTTTCCCAGTCACGAC GTTGTAAAACG	Sequencing primer.
M13 1233 reverse	AGCGGATAACAATTTCA CACAGGA	Sequencing primer.
5'MfdRID	CATACCATGGTCCGTAA CCTTGCGGAACTGC	Introduces *Nco*I site upstream of Mfd amino acid 472 for cloning into pAS vectors. Anneals to non-coding strand.
3'MfdRID	CACAGGATCCTTACGGC GTGGTTTCAAACGG	Introduces *Bam*HI site and stop codon downstream of Mfd amino acid 603 for cloning into pAS vectors. Anneals to coding strand.

Nucleic Acids Research, 2003, Vol. 31, No. 22 6409–6418
DOI: 10.1093/nar/gkg868

A DNA translocation motif in the bacterial transcription–repair coupling factor, Mfd

A. L. Chambers, A. J. Smith and N. J. Savery*

University of Bristol, Department of Biochemistry, School of Medical Sciences, University Walk, Bristol BS8 1TD, UK

Received September 9, 2003; Revised and Accepted September 30, 2003

ABSTRACT

The bacterial transcription–repair coupling factor, Mfd, is a superfamily II helicase that releases transcription elongation complexes stalled by DNA damage or other obstacles. Transcription complex displacement is an ATP-dependent reaction that is thought to involve DNA translocation without the strand separation associated with classical helicase activity. We have identified single amino acid substitutions within Mfd that disrupt the ability of Mfd to displace RNA polymerase but do not prevent ATP hydrolysis or binding to DNA. These substitutions, or deletion of the C-terminal 209 residues of Mfd, abrogate the ability of Mfd to increase the efficiency of roadblock repression *in vivo*. The substitutions fall in a region of Mfd that is homologous to the 'TRG' motif of RecG, a protein that catalyses ATP-dependent translocation of Holliday junctions. Our results define a translocation motif in Mfd and suggest that Mfd and RecG couple ATP hydrolysis to translocation of DNA in a similar manner.

INTRODUCTION

DNA is subjected to a continuous barrage of assaults from endogenous and exogenous sources, and cells employ a wide range of enzyme activities to recognise and repair DNA damage and thus ensure the integrity of the genetic material. DNA damage within genes that are being transcribed is prioritised by transcription–repair coupling pathways: lesions that occur in the transcribed strand and cause RNA polymerase (RNAP) to stall are repaired ~10-fold more rapidly than similar damage in the bulk DNA (1,2). The process is mediated by transcription–repair coupling factors that displace the stalled RNAP from the site of damage, and interact with repair proteins (3,4).

The bacterial transcription–repair coupling factor is the Mfd protein, which binds to stalled transcription elongation complexes *in vitro* and catalyses their dissociation from template DNA in an ATP-dependent manner (5). Sequence analysis indicates that Mfd contains superfamily II helicase motifs (5). These are essential for the RNAP displacement activity of Mfd (6), but Mfd does not possess conventional strand-separating helicase activity (7). In order to displace RNAP, Mfd requires ~20 bp of accessible DNA upstream of the region bound by RNAP but little or no DNA downstream of RNAP (8). Mfd is also able to 'rescue' elongation complexes that have become backtracked, allowing them to resume transcription. In backtracked complexes, the 3′ end of the nascent transcript has been extruded from the active site through backwards translocation of RNAP (8). Transcription elongation factors GreA and GreB rescue backtracked complexes by causing the extruded 3′ end of the transcript to be cleaved (9), but Mfd appears to directly reverse the backtracking process by causing the stalled RNAP to move forward (8).

These results support a model in which ATP hydrolysis by Mfd is coupled to dissociation of RNAP by translocation of double-stranded DNA (8). At stalled elongation complexes, an Mfd molecule interacts with RNAP and with the region of DNA immediately upstream of the complex. If transcription is blocked by a DNA lesion, or by a protein bound to the DNA in front of the polymerase, translocation by Mfd results in dissociation of RNAP. In the case of backtracked complexes, the translocation results in realignment of the 3′ end of the transcript and resumption of transcription.

Double-stranded DNA translocation is thought to be important for a range of processes including chromatin remodelling, DNA replication and DNA cleavage by type I restriction enzymes (10–12). One of the best characterised DNA translocases is the bacterial RecG protein, which is involved in the rescue of stalled replication forks and can catalyse branch migration of a Holliday junction structure *in vitro* in an ATP-dependent manner (13). The region of Mfd encompassing the helicase motifs shares a high degree of sequence identity with the helicase motifs of RecG (38% identity between 378 residues of *Escherichia coli* Mfd and the corresponding region of *E.coli* RecG) (5,14). The crystal structure of a *Thermatoga maritima* RecG-ADP–DNA complex (12) reveals that RecG consists of three domains: an N-terminal domain, which binds to the replication fork, and two C-terminal domains, which contain superfamily II helicase motifs. It is proposed that ATP hydrolysis by RecG results in translocation of double-stranded DNA by domains 2 and 3 (12), and it is these domains that contain the region of homology to Mfd.

Substitutions that prevent DNA translocation by RecG have recently been identified in the region immediately downstream of helicase motif VI (14). These substitutions fall in a helical

*To whom correspondence should be addressed Tel: +44 117 928 9708; Fax: +44 117 928 8274; Email: n.j.savery@bris.ac.uk

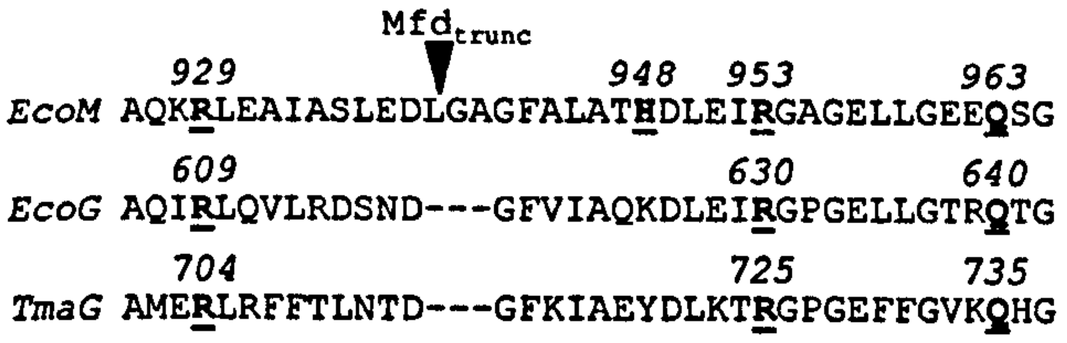


Figure 1. Alignment of the putative TRG motif of *E.coli* Mfd (*EcoM*) with the TRG motifs of *E.coli* RecG (*EcoG*) and *T.maritima* RecG (*TmaG*). The residues analysed in this work and the residues that affect TRG function in RecG are shown in bold underlined text and are numbered above each line. The C-terminal limit of the deletion mutant Mfd_{trunc} is indicated by an arrowhead.

hairpin motif and an adjacent surface-exposed loop that was not defined in the RecG crystal structure. This region was termed the TRG (translocation in RecG) motif. On the basis of Mfd:RecG sequence alignments, it was proposed that Mfd also contains a TRG motif (14) (Fig. 1). In this study, we show that the putative TRG motif in Mfd is essential for Mfd-mediated displacement of RNAP, but is not essential for ATP-dependent DNA binding. We also show that Mfd increases the efficiency of roadblock repression of transcription *in vivo*, and that residues within the TRG motif are required for this function.

MATERIALS AND METHODS

Strains and plasmids

The *mfd*$^+$ *E.coli* K-12 strain AB1157 was obtained from the *E.coli* Genetic Stock Centre, Yale University. UNCNOMFD is an *mfd*$^-$ derivative of AB1157 in which a 2317 bp NsiI–NsiI fragment of the chromosomal *mfd* gene has been replaced by a cassette encoding kanamycin resistance (6). Standard methods for isolation and manipulation of DNA fragments were used throughout (15). All constructs were confirmed by sequencing. Details of the primers used for directed and random mutagenesis are available on request.

The roadblock repression reporter vector pRCB-CAT1 (pRK2 *ori*, Tet^r) was constructed in several steps, as follows. The HindIII cassette carrying the chloramphenicol acetyltransferase gene (*cat*) from pCM7 (Pharmacia) was cloned into the HindIII site of pSR*galPcon6* [a pBR322 derivative carrying the constitutive *galPcon6* promoter (16) upstream of a *λoop* transcription terminator]. A self-complementary oligonucleotide containing the optimal *lac* operator sequence 5′-GAATTGTGAGCGCTCACAATTC-3′ (17) was cloned into the HindIII sites on either side of the *cat* gene to generate plasmid pCSB1. The NheI–HindIII *lacZYA* fragment was removed from the broad-host range *lacZ* reporter plasmid pRW50 (18), blunt ends were generated using Klenow polymerase, and the vector was religated to create plasmid pRCB. The EcoRI–BamHI fragment of pCSB1, which carried the *galpcon6* promoter, the *cat* gene flanked by *lac* operators, and the transcription terminator, was inserted between the EcoRI and BamHI sites of pRCB to generate pRCB-CAT1.

The Mfd expression plasmids pETMfd and pETTrunc carry the *mfd* gene (or a derivative) cloned into the backbone of pET21a (Novagen). Neither pETMfd nor pETTrunc contain the T7 promoter of pET21a, but both retain the *lacI*^{Q} gene that encodes the Lac repressor protein necessary for repression of the *cat* gene in pRCB-CAT1. Expression of Mfd from pETMfd and pETTrunc is controlled by the wild-type *mfd* promoter (19), and the Mfd proteins are not histidine tagged. pETMfd was constructed by ligating an EcoRI–SphI fragment of pMFD19 (5) containing the *mfd* gene and promoter with an EcoRI–SphI fragment of pET21a containing the *ori*, *bla* and *lacI*^{Q} genes. Derivatives of pETMfd encoding Mfd with RA929, RA953 and QA963 substitutions were constructed by site-directed mutagenesis (15). pETMfd encoding Mfd with an HL948 substitution was isolated from a library of pETMfd derivatives in which codons 913–995 of *mfd* had been randomly mutated by error-prone PCR using Taq DNA polymerase (20).

pETTrunc encodes residues 1–939 of Mfd (Mfd_{trunc}) and was constructed in several stages, as follows. A HindIII–StyI PCR product carrying the λoop transcription terminator from pSR*lacUV5* (21) was inserted between the HindIII and StyI sites of pET21a to create pET21term. A stop codon and a HindIII site were introduced downstream of *mfd* codon 939 by PCR, using pMFD19 as a template. The resulting PCR product, which carried the *mfd* promoter and truncated *mfd* gene, was cut with BamHI and HindIII and ligated with pET21term that had been cut with BglII and HindIII, generating pETTrunc.

CAT assays

AB1157 or UNCNOMFD cells transformed with pRCB-CAT1 and pETMfd, pETTrunc or a derivative were grown at 37°C in M9 medium containing 80 μg/ml ampicillin and 20 μg/ml tetracycline (plus 25 μg/ml kanamycin for UNCNOMFD). Where indicated, 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture prior to growth. At OD_{600} ~0.5, 1.5 ml of culture was harvested by centrifugation and the cells were lysed with 150 μl of Bugbuster (Novagen) according to the manufacturer's protocol. The protein concentrations in the cell extracts were established using the Bradford assay. The extracts were snap frozen and stored at –80°C. CAT activity was quantified using the Quan-T-CAT assay kit (Amersham) according to the manufacturer's protocol. CAT activities are expressed as nmol of chloramphenicol acetylated/min/mg of protein.

Proteins

Escherichia coli RNAP holoenzyme was purchased from Epicentre Technologies (Madison, WI). Wild-type and mutant Mfd proteins were purified from UNCNOMFD cells transformed with the appropriate pETMfd derivative. Cells were grown at 37°C in 1 l of 2× YT medium containing 80 μg/ml ampicillin and 25 μg/ml kanamycin. At an OD_{600} of ~1.7, the culture was harvested by centrifugation, the pellet was resuspended in 20 ml of lysis buffer [50 mM Tris pH 7.5, 5% glycerol (v/v), 120 mM NaCl, 10 mM EDTA, 5 mM dithiothreitol (DTT)] and the cells were lysed by sonication. Mfd was purified from the lysate using a 5 ml HiTrap Blue Sepharose column, a 5 ml HiTrap heparin column and a 1 ml Mono Q column (Amersham). Protein was eluted from each column with a 0.05–1.0 M KCl gradient in TGED buffer [10 mM Tris pH 7.5, 5% glycerol (v/v), 1 mM EDTA, 1 mM DTT], and fractions containing Mfd were diluted 10-fold in TGED before being loaded onto the next column. Purified

protein was dialysed against storage buffer [10 mM Tris pH 8.0, 50% glycerol (v/v), 1 mM EDTA, 2 mM DTT, 200 mM KCl] and stored at –20°C. Protein concentrations were determined using the Bradford assay.

In vitro analysis of RNAP displacement

Stalled transcription elongation complexes were formed by nucleotide starvation on a 529 bp RsaI–SmaI fragment of plasmid pAR1707 (22). This fragment encodes a transcript in which UTP must be incorporated at +2 and +21. Transcription elongation complexes stalled at +20 could therefore be formed by using the dinucleotide ApU to initiate transcription and excluding UTP from the transcription reaction. Complexes were analysed by electrophoretic mobility shift assays (EMSAs) using 4.5% acrylamide/1× TBE gels run at 4°C, and radiolabelled bands were detected using a phosphor screen and quantified using Imagequant software (Molecular Dynamics).

For experiments to determine the time course of Mfd-mediated displacement of RNAP, the fragment of pAR1707 was end labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP (15). Initiation complexes were formed by incubating 0.4 nM labelled fragment with 5 nM RNAP holoenzyme in repair assay buffer [40 mM HEPES, pH 8.0, 100 mM KCl, 8 mM $MgCl_2$, 4% glycerol (v/v), 5 mM DTT, 100 µg/ml bovine serum albumin (BSA)] for 15 min at 37°C. Unstable complexes were removed by the addition of 10 µg/ml heparin, then ApU, ATP, GTP and CTP were added at final concentrations of 80 µM, 1.7 mM, 8 µM and 8 µM, respectively, to initiate transcription. The reaction mixtures were then incubated at 37°C for 15 min. Mfd was added at a final concentration of 250 nM and, after the indicated incubation periods at 37°C, aliquots were withdrawn and loaded onto an acrylamide gel under tension.

In experiments to analyse binding of mutant Mfd to transcription elongation complexes, either the fragment of pAR1707 was end-labelled as above, or the nascent transcripts were labelled by incorporation of [α-^{32}P]CTP. For reactions using labelled template, initiation complexes were formed as described above. Where indicated, ApU, ATP, CTP and GTP were added at final concentrations of 80 µM, 0.8 mM, 8 µM and 8 µM, respectively. The reaction mixtures were incubated at 37°C for 15 min and then 5.5 µg/ml rifampicin was added. Mfd was added at a final concentration of 400 nM and samples were incubated at 37°C for 15 min before being analysed by gel electrophoresis. For reactions in which the nascent transcripts were labelled, initiation complexes were formed by incubating 2 nM unlabelled pAR1707 fragment with 10 nM RNAP holoenzyme in repair assay buffer at 37°C for 15 min. Rifampicin and then ApU, ATP and GTP were added at final concentrations of 20 µg/ml, 100 µM, 2 mM and 10 µM, respectively. Unlabelled CTP was added at a final concentration of 2 µM, together with 5 µCi of [α-^{32}P]CTP per reaction. The reaction mixtures were incubated at 37°C for 15 min. Mfd was then added at a final concentration of 400 nM and the samples incubated at 37°C for 15 min before being analysed by gel electrophoresis.

DNA binding assays

Binding of Mfd to DNA was assayed by EMSAs (23). A 248 bp EcoRI–BamHI fragment from pSR*lacUV5* (21) was end labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP (15). Reactions contained 100 nM purified Mfd and 0.4 nM labelled fragment in 15 µl of repair assay buffer. Where indicated, the reactions also contained 2 mM ATP or ATPγS. Reactions were incubated at 37°C for 25 min and analysed by electrophoresis at 4°C through 5% acrylamide/1× TAE/8 mM magnesium acetate gels. Bands were detected using a phosphor screen.

ATPase assay

Rates of ATP hydrolysis were determined by measuring the release of ^{32}P-labelled inorganic phosphate from [γ-^{32}P]ATP, essentially as described by Tombline and Fishel (24). Reaction mixtures contained 1 µM Mfd and 2 mM ATP in repair buffer at 37°C.

RESULTS

A roadblock repression assay for Mfd function

Proteins bound to DNA in the path of a transcription elongation complex can cause RNAP to pause or stall, thus decreasing transcription of sequences downstream of the protein-binding site. Regulated formation of protein roadblocks is involved in carbon catabolite repression of the *hut* and *gnt* operons in *Bacillus subtilis* (25,26), and catabolite repression of these operons is partially relieved in *mfd*$^-$ cells (27). It has been proposed that the Mfd dependence of roadblock repression in *B.subtilis* is the result of Mfd-mediated displacement of RNAP paused at roadblock sites (27). We established an *in vivo* assay to test this model in *E.coli* and to provide an Mfd-dependent phenotype that could be readily monitored.

The rationale for the assay is shown in Figure 2. Lac repressor–operator interactions are sufficient to cause roadblock repression both *in vivo* and *in vitro*, although roadblock repression is less efficient than repression of the *lacP1* promoter by promoter occlusion (28–31). We constructed a reporter vector, pRCB-CAT1, in which roadblock repression by Lac repressor protein controlled expression of the *cat* gene. In order to maximise the level of roadblock repression, we placed one optimal *lac* operator (17) upstream of the *cat* gene, and a second optimal *lac* operator downstream of the *cat* gene. The two operators were separated by 815 bp (centre-to-centre), a distance at which simultaneous binding of the two operators by a single Lac repressor tetramer would be expected to increase occupancy of the upstream operator ~2-fold (32) without introducing a second roadblock between the promoter and the reporter gene. Expression of the *cat* gene was driven by a constitutive derivative of the *galP1* promoter, located 64 bp upstream of the centre of the upstream *lac* operator. Lac repressor bound to the upstream *lac* operator was expected to block the progress of RNAP and decrease transcription of the *cat* gene. If Mfd displaces transcription complexes stalled by protein roadblocks *in vivo*, as it does *in vitro* (7), it should increase the likelihood that the stalled RNAP dissociates from the DNA before Lac repressor does, decreasing transcription of the *cat* gene still further. The Lac repressor roadblock should therefore repress transcription more effectively when Mfd is present than when Mfd is absent.

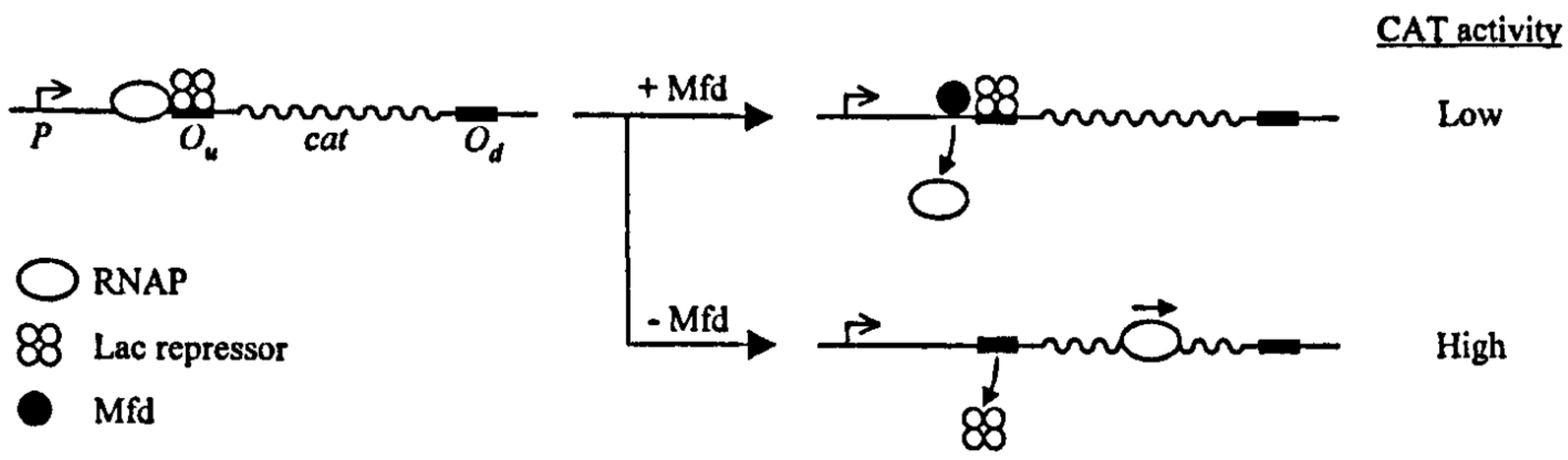


Figure 2. Roadblock reporter assay for Mfd function. Roadblock repression reporter vector pRCB-CAT1 contains the *cat* gene flanked on either side by *lac* operators (O_u and O_d). The *cat* gene is transcribed from the constitutive promoter, *galpcon6* (*P*). A Lac repressor tetramer bound to the upstream operator (O_u) will cause RNAP to stall upstream of the *cat* gene, and the downstream operator (O_d) enhances occupancy of O_u. Mfd increases the likelihood that RNAP will dissociate before Lac repressor does, and so the *cat* gene is transcribed less frequently when Mfd is present.

Table 1. Effect of Mfd on roadblock repression *in vivo*

Mfd	+IPTG (induced)	–IPTG (repressed)	Repression (+IPTG/–IPTG)
None	24.0 ± 2.1	2.5 ± 0.5	9.5-fold
Wild-type	15.2 ± 2.0	0.3 ± 0.1	61-fold
Mfd_{trunc}	20.2 ± 2.3	3.1 ± 0.5	6.5-fold

UNCNOMFD (*mfd*$^-$) cells were transformed with the roadblock repression reporter vector pRCB-CAT1 and with plasmids encoding the Mfd derivatives indicated. Wild-type Mfd was expressed from pETMfd, Mfd_{trunc} was expressed from pETTrunc, and pET21a was used as a control plasmid that did not encode Mfd. Cultures were grown to OD_{600} ~0.5 in M9 medium + antibiotics ± 0.5 mM IPTG as indicated. CAT activities shown (with SD) are the average of three independent experiments and are expressed in units of nmol of chloramphenicol acetylated/min/mg of protein.

To determine the effectiveness of this roadblock repression system in the absence of Mfd, an *mfd*$^-$ strain, UNCNOMFD, was transformed with the roadblock repression reporter vector and a plasmid carrying the *lacI*^{Q} gene encoding Lac repressor. Cultures of this strain were grown to mid-exponential phase in the presence or absence of IPTG, and CAT activities were measured (Table 1). The CAT activity in the absence of IPTG (when Lac repressor should be bound to operator DNA) was 9.5-fold lower than that in the presence of IPTG (when Lac repressor should not be bound to operator DNA). Lac repressor bound downstream of the transcription start site therefore repressed transcription of the *cat* gene 9.5-fold in this system, presumably by acting as a block to transcription elongation.

To determine the effect of Mfd on roadblock repression in this system, the experiment was repeated using UNCNOMFD (*mfd*$^-$) cells transformed with the roadblock repression reporter vector and a plasmid carrying the wild-type *mfd* gene in addition to *lacI*^{Q} (Table 1). In these cells, the CAT activity in the absence of IPTG was 61-fold lower than that in its presence. Comparison with the results obtained in cells that lacked Mfd reveals that Mfd decreased the induced (+IPTG) level of CAT expression only slightly, and caused a much greater decrease in the repressed (–IPTG) level of CAT expression. We conclude that Mfd substantially increases the efficiency with which a Lac repressor roadblock represses transcription in *E.coli*.

The effect of Mfd on the repressed level of CAT expression was sufficient to result in differential growth of *mfd*$^+$ and *mfd*$^-$ strains on agar plates containing chloramphenicol (data not shown). UNCNOMFD (*mfd*$^-$) cells transformed with the roadblock repression reporter vector and a control plasmid (pET21a) that did not encode Mfd grew readily on M9 minimal medium agar containing 5 µg/ml chloramphenicol in the absence of IPTG. Identical cells transformed with a plasmid that encoded wild-type Mfd (pETMfd) were unable to grow under the same conditions. A chloramphenicol-sensitive phenotype in this assay is therefore an indication of Mfd activity. Mutants that are defective in the pathway by which Mfd affects roadblock repression should confer a chloramphenicol-resistant phenotype and thus be readily selectable. Several mutants with such a phenotype are described below.

The C-terminal region of Mfd is required for optimal roadblock repression

A truncated version of Mfd (Mfd_{trunc}) lacking the C-terminal 209 amino acids has previously been shown to be specifically defective in RNAP displacement, but to retain the ability to bind to both RNAP and DNA (6). If the effect of Mfd on roadblock repression is dependent on its ability to dissociate stalled transcription elongation complexes, Mfd_{trunc} should be unable to enhance roadblock repression *in vivo*. We tested this prediction using the CAT reporter assay described above. UNCNOMFD (*mfd*$^-$) cells were transformed with the roadblock repression reporter vector and a plasmid encoding Mfd_{trunc} and *lacI*^{Q}. The level of *cat* gene expression was monitored by testing the ability of the transformants to grow on M9 agar plates containing 5 µg/ml chloramphenicol and by assaying CAT activity in liquid cultures. The transformants were resistant to 5 µg/ml chloramphenicol (data not shown), and the CAT activity in the absence of IPTG was 6.5-fold lower than that in its presence (Table 1). Both the

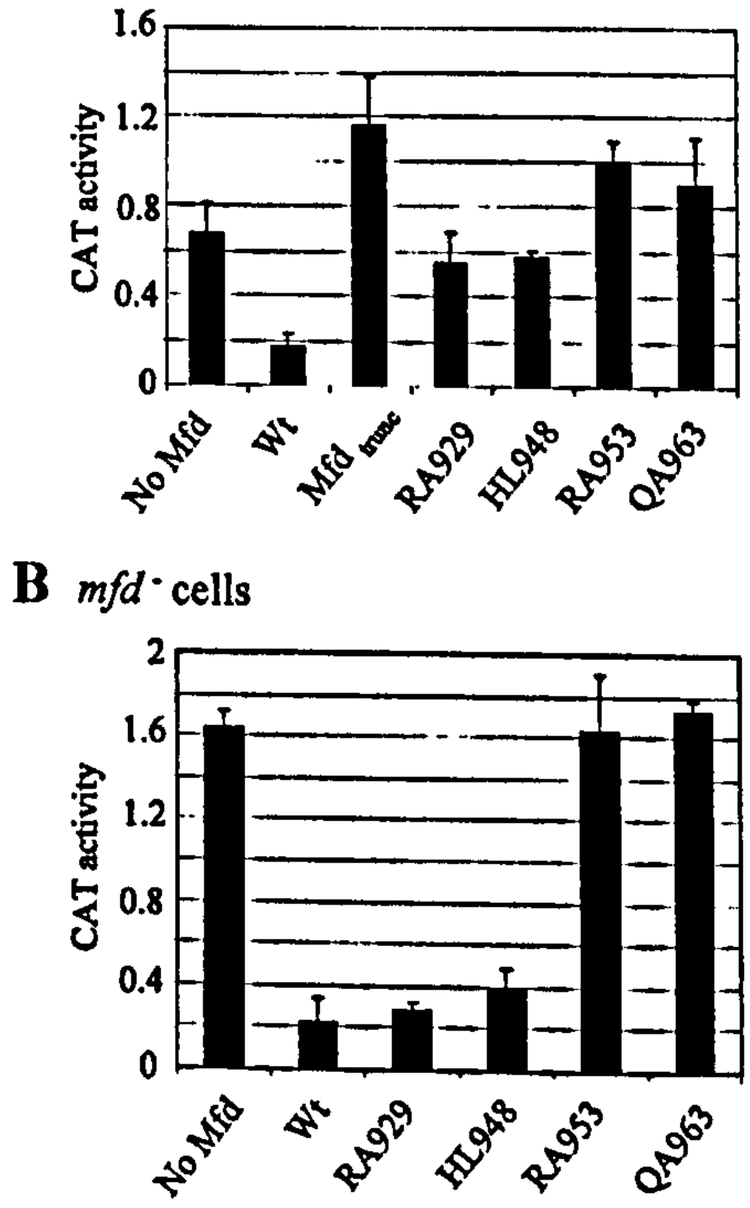


Figure 3. Effect of single amino acid substitutions in Mfd on roadblock repression *in vivo*. (A) AB1157 (*mfd*$^+$) or (B) UNCNOMFD (*mfd*$^-$) cells were transformed with the roadblock repression reporter vector pRCB-CAT1 and with plasmids encoding the Mfd derivatives indicated. Wild-type Mfd (Wt) and full-length Mfd proteins carrying single amino acid substitutions were expressed from plasmid pETMfd, or derivatives. Mfd_{trunc} was expressed from plasmid pETTrunc. Control cells that lacked plasmid-encoded *mfd* (No Mfd) contained plasmid pET21a. Cultures were grown to OD_{600} ~0.5 in the absence of IPTG in M9 medium containing the appropriate antibiotics. CAT activities shown (with SD) are the average of three independent experiments and are expressed in units of nmol of chloramphenicol acetylated/min/mg of protein.

chloramphenicol resistance phenotype and the efficiency of roadblock repression in these cells were therefore comparable with results obtained with cells that lacked Mfd, indicating that Mfd_{trunc} did not increase the efficiency of roadblock repression *in vivo*.

The effect of Mfd_{trunc} expression was also determined in cells carrying an intact chromosomal *mfd* gene (Fig. 3A). The repressed level of CAT activity in *mfd*$^+$ cells transformed with a plasmid that encoded Mfd_{trunc} was higher than in those transformed with a control plasmid that did not encode Mfd. In control experiments, the repressed level of CAT activity in *mfd*$^+$ cells transformed with a plasmid that encoded wild-type Mfd was ~3-fold lower than in those transformed with the control plasmid. The ability of Mfd_{trunc} to exert a *trans*-dominant effect over wild-type Mfd confirmed that Mfd_{trunc} was expressed, and suggests that despite being unable to dissociate the RNAP, it was able to compete with wild-type Mfd for binding sites on the stalled RNAP complexes. The increased repression in cells that contained plasmid-borne *mfd* genes in addition to the chromosomal gene indicates that wild-type levels of Mfd were not saturating.

The results obtained with Mfd_{trunc} confirm that the ability of Mfd to increase the efficiency of roadblock repression is dependent on its ability to displace RNAP from DNA. Expression of genes that are controlled via roadblock repression should therefore be sensitive to mutations that abolish the ability of Mfd to displace RNAP from DNA. Within the *mfd* gene, such mutations could affect Mfd–RNAP binding, Mfd–DNA binding, the ATPase activity of Mfd, or the mechanism by which ATP hydrolysis is linked to RNAP displacement.

Substitutions in the TRG motif affect Mfd function *in vivo*

To determine whether the putative TRG motif plays a functional role in RNAP displacement by Mfd, we generated Mfd derivatives carrying single amino acid substitutions at positions that correspond to key residues in the TRG motif of *E.coli* RecG (14) (Fig. 1). Site-directed mutagenesis was used to introduce single alanine substitutions at positions R929, R953 and Q963 of Mfd. In addition, we used error-prone PCR to introduce random mutations into codons 913–995 of *mfd* (a region encompassing the putative TRG motif). The library of mutants was introduced into UNCNOMFD (*mfd*$^-$) cells carrying the roadblock repression reporter vector, and transformants were screened for mutants that were defective in RNAP displacement, as judged by their ability to grow on M9 minimal agar containing 5 μg/ml chloramphenicol. A non-saturating screen yielded a chloramphenicol-resistant mutant encoding a leucine substitution at position H948 of Mfd.

The ability of the mutant Mfd proteins to displace stalled transcription complexes *in vivo* was assayed by monitoring their effect on roadblock repression using the CAT reporter system described above. UNCNOMFD (*mfd*$^-$) and AB1157 (*mfd*$^+$) cells were transformed with the roadblock repression reporter vector and pETMfd derivatives encoding Mfd RA929, HL948, RA953 and QA963. The effect of the Mfd derivatives on roadblock repression was monitored by directly assaying CAT activity in liquid cultures (Fig. 3) and by testing the ability of the transformants to grow on M9 agar plates containing 5 μg/ml chloramphenicol (data not shown).

UNCNOMFD (*mfd*$^-$) cells transformed with the roadblock repression reporter vector and with plasmids encoding Mfd RA953 and QA963 grew on M9 agar plates containing 5 μg/ml chloramphenicol. The repressed level of CAT activity measured during growth of these cells in liquid culture was similar to that observed in the absence of Mfd (Fig. 3B). Both Mfd RA953 and Mfd QA963 could be purified from UNCNOMFD transformants with a yield similar to wild-type (data not shown), demonstrating that they were stably expressed in this system. These results indicate that Mfd RA953 and Mfd QA963 did not affect the efficiency of roadblock repression *in vivo* and that the substitutions have affected a function of Mfd that is essential for displacement of RNAP. Mfd RA953 conferred a *trans*-dominant increase in CAT activity in AB1157 (*mfd*$^+$) cells (Fig. 3A), suggesting that, like Mfd_{trunc}, the protein was able to compete with wild-type Mfd for binding sites on the stalled RNAP complexes.

UNCNOMFD (*mfd*$^-$) cells transformed with the roadblock repression reporter vector and with a plasmid encoding Mfd RA929 did not grow on M9 agar plates containing 5 μg/ml chloramphenicol, and the repressed level of CAT activity measured during growth of these cells in liquid culture was similar to that observed with cells transformed with a plasmid

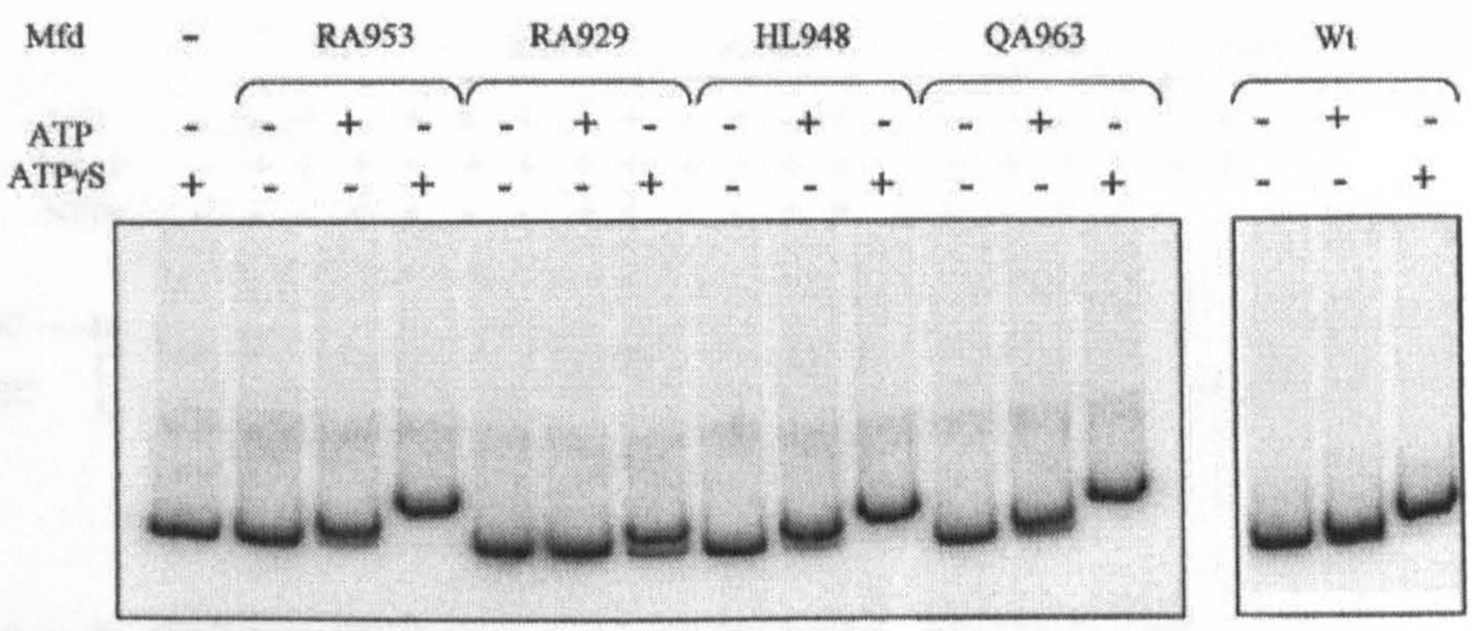


Figure 4. Effect of single amino acid substitutions in Mfd on DNA binding *in vitro*. A 100 nM concentration of wild-type Mfd (Wt), or Mfd carrying the substitutions indicated, was incubated for 25 min at 37°C with 0.4 nM end-labelled EcoRI–BamHI DNA fragment ±2 mM ATP or ATPγS, as indicated. The complexes were analysed using a 5% acrylamide gel.

encoding wild-type Mfd (Fig. 3B). These results indicate that Mfd RA929 was able to increase the efficiency of roadblock repression *in vivo* and must therefore be capable of displacing stalled transcription elongation complexes.

UNCNOMFD (*mfd*⁻) cells transformed with the roadblock repression reporter vector and with a plasmid encoding Mfd HL948 grew on M9 agar plates containing 5 μg/ml chloramphenicol, but the repressed level of CAT activity measured during growth of these cells in liquid culture was similar to that observed with cells transformed with a plasmid encoding wild-type Mfd (Fig. 3B). These results are apparently contradictory, as growth on plates containing 5 μg/ml chloramphenicol requires a higher level of CAT expression than occurs in cells containing wild-type Mfd. The discrepancy presumably reflects a difference in physiology between cells grown in liquid media and cells grown on solid media. The result of the CAT assay indicates that Mfd HL948 was able to increase the efficiency of roadblock repression *in vivo* and must therefore be capable of displacing stalled transcription elongation complexes, at least under certain conditions. As *in vitro* analysis revealed that the RNAP displacement activity of Mfd HL948 was functional but impaired (see below), it appears that the chloramphenicol resistance phenotype on solid media is a more sensitive assay of Mfd function than CAT activity in liquid culture.

DNA and ATP binding by purified mutant Mfd proteins

The Mfd derivatives carrying substitutions in the TRG motif were purified, and their properties analysed *in vitro*. Mfd has been shown to possess non-specific ATP-dependent DNA binding activity (5,6), and we tested the DNA binding properties of the purified mutant Mfd proteins in the presence of ATP or ATPγS (a poorly hydrolysable ATP analogue) and in the absence of nucleotide (Fig. 4). In accordance with previously published work (6), wild-type Mfd only formed a stable complex in the presence of ATPγS. Mfd RA953, HL948 and QA963 also formed stable complexes in the presence of ATPγS but, in contrast to wild-type Mfd, these mutant proteins formed a significant amount of stable complex in the presence of ATP. This ATP-induced complex had a higher electrophoretic mobility than the complex formed in the presence of ATPγS, suggesting that the two complexes differ in character. Mfd RA929 did not form a stable complex in the presence of ATP, and the complex formed by Mfd RA929 in the presence of ATPγS resembled that formed by the other mutants in the presence of ATP.

The observation that all four Mfd derivatives carrying substitutions in the TRG motif bound DNA stably in the presence of ATPγS, but not in the absence of nucleotides, confirmed that they were able to bind ATP and suggested that the substitutions did not grossly affect the overall fold of the protein. We assayed the ATPase activity of our purified Mfd proteins at saturating ATP concentration in the absence of DNA. Under these conditions, all four of the Mfd derivatives carrying substitutions in the TRG motif hydrolysed ATP at the same rate, within the error of the experiment (data not shown). Wild-type Mfd presumably does not bind DNA stably in the presence of ATP because hydrolysis of ATP causes Mfd to be lost from the fragment. This might occur either by a simple conformational change induced by ATP hydrolysis or product release, or through ATP-driven DNA translocation. The ability of Mfd RA953, HL948, and QA963 to form stable complexes in the presence of ATP might therefore result either from defects in the ATP hydrolysis cycle that were too subtle to be detected in our ATPase assay, or an inability to couple ATP hydrolysis to DNA translocation or release.

Substitutions in the TRG motif disrupt Mfd-mediated displacement of RNAP *in vitro*

The ability of purified mutant Mfd proteins to displace stalled transcription complexes *in vitro* was assayed by EMSAs (Fig. 5). Plasmid pAR1707 contains the T7 A1 promoter, and elongation complexes initiating from this promoter can be stalled at +20 by excluding UTP from a reaction mixture that contains all other components necessary for transcription (22,33). Experiments using ^{32}P-labelled template DNA were used to detect transcription initiation complexes, transcription elongation complexes and DNA to which RNAP had not bound (Fig. 5A), and experiments in which ^{32}P-labelled nucleotides were incorporated into the nascent transcript were used to detect transcription elongation complexes specifically (Fig. 5B). Initiation complexes and stalled elongation complexes were distinguished on the basis of their differing mobility during acrylamide gel electrophoresis (e.g. compare

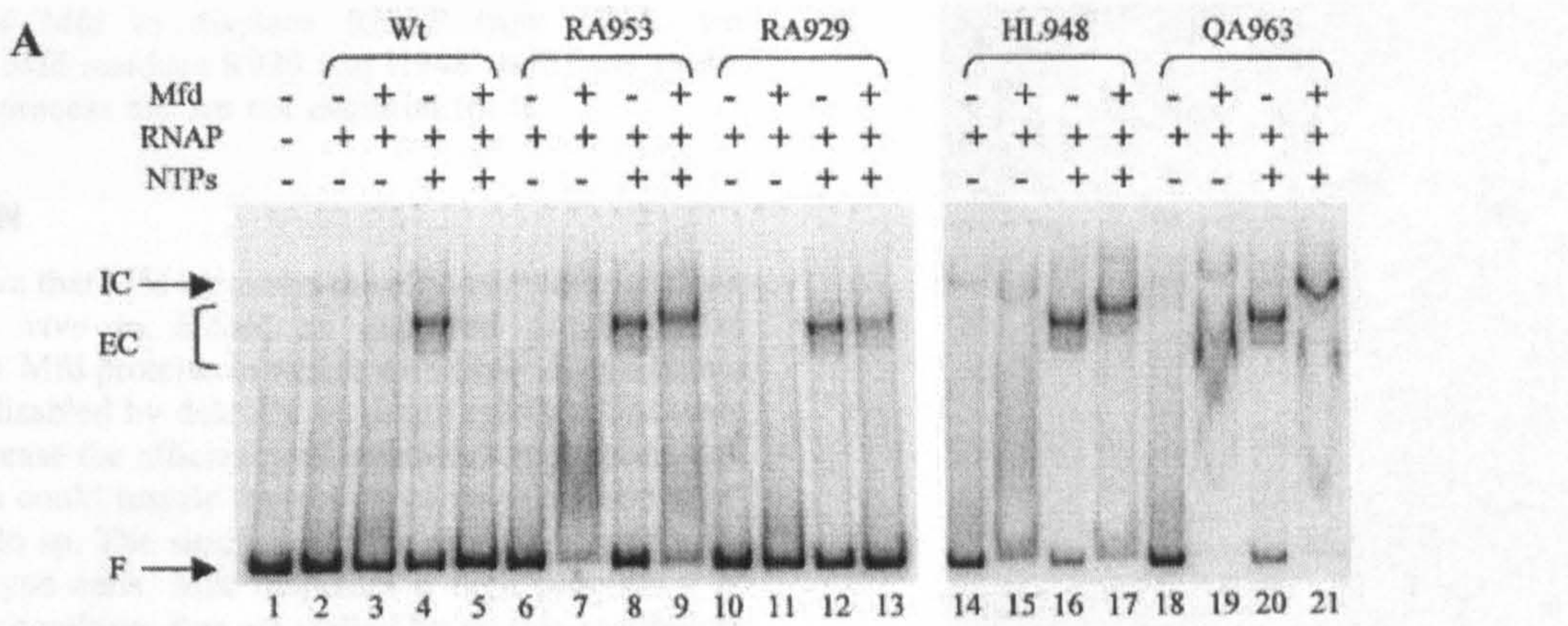


Figure 5. Effect of single amino acid substitutions in Mfd on Mfd-mediated displacement of RNAP *in vitro*. Transcription complexes were formed on an RsaI–SmaI fragment of plasmid pAR1707 and analysed by EMSA. Initiation complexes were formed by the addition of RNAP holoenzyme (RNAP) alone, and elongation complexes stalled at +20 were formed by the addition of RNAP and a mixture of ApU, ATP, CTP and GTP (NTPs). Purified wild-type Mfd (Wt) or Mfd carrying the single amino acid substitutions shown were added where indicated. Bands corresponding to transcription initiation complexes (IC), elongation complexes (EC) and unbound fragment (F) are indicated. (**A**) Image of an EMSA gel used to analyse interaction of Mfd with transcription complexes using template DNA end-labelled with ^{32}P. (**B**) Image of an EMSA gel used to analyse binding of Mfd to transcription elongation complexes in which the nascent transcript was labelled by the incorporation of [α-^{32}P]CTP. (**C**) Time course of displacement of stalled elongation complexes at 37°C by wild-type Mfd (open circles), Mfd RA929 (open triangles), Mfd HL948 (filled triangles), Mfd RA953 (open squares), Mfd QA963 (filled squares) and in the absence of Mfd (filled circles). Values are averages obtained by quantitative analysis of at least three independent EMSA experiments and are expressed as a percentage of the amount of elongation complex at $t = 0$.

lanes 2 and 4 in Fig. 5A). Initiation complexes migrated as a single band, and were less stable than elongation complexes under our electrophoresis conditions. Elongation complexes migrated faster than the initiation complex, and several minor species were detected in addition to the major band (e.g. lane 4 in Fig. 5A and lane 2 in Fig. 5B). Previous studies using this template indicate that the fainter, faster moving bands correspond to elongation complexes that contain transcripts shorter than 20 nucleotides (22,33).

As described previously (5,6,8), addition of wild-type Mfd to a reaction mixture containing stalled elongation complexes resulted in the complete dissociation of all elongation complexes (Fig. 5A, compare lanes 4 and 5; Fig. 5B, compare lanes 1 and 2). Analysis of samples taken at timed intervals after the addition of Mfd revealed that >90% of the elongation complexes were displaced from the DNA within 5 min (Fig. 5C). In contrast, Mfd RA953 and Mfd QA963 were unable to displace any elongation complexes, even after 20 min incubation (Fig. 5C). However, in reactions containing Mfd RA953 or Mfd QA963, the transcription elongation complex band was 'supershifted' (i.e. its mobility was decreased; compare lanes 2, 3 and 6 in Fig. 5B; supershifting by QA963 is also evident in Fig. 5A). We conclude that alanine substitution of R953 or R963 abolishes the ability of Mfd to displace RNAP from DNA, but that the mutant proteins retain the ability to bind to stalled elongation complexes. As both Mfd RA953 and QA963 bound naked DNA non-specifically under the conditions of these assays (note the shifting of the fragment band in Fig. 5A, and see Fig. 4), it is not clear whether the supershifted elongation complexes involved Mfd–RNAP–DNA interactions or solely Mfd–DNA interactions.

Mfd RA929 and Mfd HL948 were able to displace stalled elongation complexes, but did so at a rate that was considerably slower than wild-type Mfd (Fig. 5C). In samples taken before the elongation complex was displaced, Mfd HL948 caused a supershift of the elongation complex band (Fig. 5A, lane 17, and Fig. 5B, lane 5) but no supershifted intermediate was observed in reactions containing Mfd RA929. As these substitutions impaired, rather than abolished,

the ability of Mfd to displace RNAP from DNA, we conclude that Mfd residues R929 and H948 contribute to the displacement process but are not essential for it.

DISCUSSION

We have shown that Mfd increases the efficiency of roadblock repression *in vivo* in *E.coli*, as observed previously in *B.subtilis* (27). Mfd proteins in which the RNAP displacement activity was disabled by deletion or single substitutions were unable to increase the efficiency of roadblock repression, and in some cases could impair the ability of co-expressed wild-type Mfd to do so. The simplest explanation of these data is that in wild-type cells, Mfd displaces a high proportion of transcription complexes that are stalled by protein roadblocks before the repressor protein dissociates from the DNA. In addition to its role in roadblock repression, Mfd is able to revive backtracked elongation complexes (8), and is responsible for RNAP release during transcription termination by the phage λ Nun protein (34). It is increasingly apparent that Mfd plays a more general role in the regulation of transcription elongation than simply mediating transcription-coupled repair.

Mfd couples ATP hydrolysis to movement of DNA–RNAP complexes, and RecG couples ATP hydrolysis to movement of specific DNA structures. Both proteins therefore require a 'motor' domain. Sequence alignment of the two proteins revealed that, in addition to conserved superfamily II helicase motifs, they share an area of high sequence identity that includes a helical hairpin motif termed the TRG motif (14). As mutational analysis indicated that residues within this motif were important for DNA translocation by RecG (14), it was likely that the TRG motif would also be important for the function of Mfd. We tested this hypothesis by making single amino acid substitutions at the residues within Mfd that correspond to the key residues in the TRG motif of RecG (Fig. 1). All of these substitutions caused defects in Mfd.

Alanine substitutions of R953 and Q963 abolished the ability of Mfd to displace stalled transcription elongation complexes *in vitro*, and produced Mfd derivatives that were unable to complement an *mfd* deletion mutant *in vivo*. These residues therefore appear to play essential roles in the mechanism by which Mfd dissociates RNAP from DNA. Substitution of R929 or H948 impaired the RNAP displacement activity of Mfd *in vitro*, but proteins containing these substitutions retained sufficient activity to complement an *mfd* deletion mutant *in vivo*. The DNA and ATP binding activities of all four mutants appeared to be intact. The 'hierachy' of effects of TRG substitutions in Mfd mirrored that observed in RecG: substitutions of equivalent residues caused defects of equivalent magnitude in both proteins. In Mfd, alanine substitutions of R953 and Q963 caused greater defects than substitution of R929 *in vivo* and *in vitro*. In *E.coli* RecG, substitutions of R630 (equivalent to Mfd R953) and Q640 (equivalent to Mfd Q963) had a greater effect on the *in vivo* role of RecG in UV tolerance than substitutions of R609 (equivalent to Mfd R929), and substitutions of R630 caused greater defects in RecG helicase and ATPase activity *in vitro* than substitutions of R609 (the effects of RecG Q640 substitutions *in vitro* have not been reported) (14). We conclude that the TRG motif of Mfd is essential for Mfd-mediated displacement of stalled RNAP, and that Mfd couples ATP hydrolysis to RNAP displacement via a mechanism very similar to that used by RecG to drive migration of Holliday junctions. Our results therefore support a model for Mfd action in which dissociation of RNAP is dependent on translocation of double-stranded DNA by Mfd.

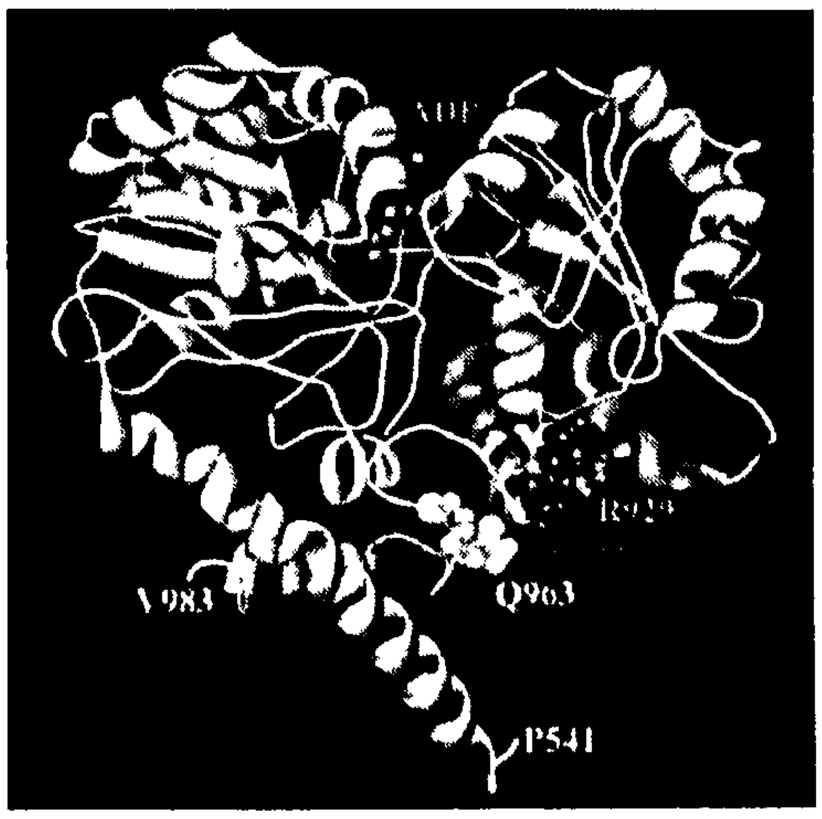


Figure 6. Model of residues 541–983 of Mfd. The model is based on the coordinates of the *T.maritima* RecG-ADP–DNA complex [accession code 1GM5; (12)] and was constructed by homology modelling using ClustalW and InsightII 97 (Accelrys). The image was produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco. The orange ribbon denotes the helical hairpin of the TRG motif, the yellow ribbon denotes the surface-exposed loop of the TRG motif, and the green ribbon denotes helicase motif VI. The residues analysed in this study are labelled, with the exception of H948 (white spheres), which is partially obscured by R953.

The primary sequences of the helicase motifs and TRG motif of Mfd and RecG are highly homologous, and we constructed a model of residues 541–983 of Mfd based on the crystal structure of residues 309–755 of *T.maritima* RecG (Fig. 6). Within this model, Mfd residues R929 and R953 are closely juxtaposed at the base of a helical hairpin, and Mfd residue Q963 falls in a surface-exposed loop that was not resolved in the RecG-ADP–DNA crystal structure. Mfd H948, which is not conserved in RecG or in Mfd sequences from different species, falls within the helical hairpin bounded by R929 and R953, and is adjacent to D949, a highly conserved residue that is implicated in TRG function in RecG (14). R953 is suitably located to form hydrogen bonds with D889 (a residue at the N-terminal end of a helix that extends into helicase motif VI) and with D949.

Mahdi and co-workers (14) suggested a plausible model for the generation of ATP-driven conformational change in the TRG motif, based on the crystal structure of RecG (12). Despite differences in the size of the TRG helical hairpin, and some differences in the residues involved, this model appears to be equally applicable to Mfd. The model, adapted for Mfd, can be summarised as follows: an arginine residue at the C-terminal end of helicase motif VI interacts with the γ phosphate of ATP. The orientation of the helix containing helicase motif VI changes during the cycle of ATP binding, hydrolysis and release, and consequently so does the pattern of hydrogen bonds made between D889 and the arginines within

the ability of Mfd to displace RNAP from DNA, we conclude that Mfd residues R929 and H948 contribute to the displacement process but are not essential for it.

DISCUSSION

We have shown that Mfd increases the efficiency of roadblock repression *in vivo* in *E.coli*, as observed previously in *B.subtilis* (27). Mfd proteins in which the RNAP displacement activity was disabled by deletion or single substitutions were unable to increase the efficiency of roadblock repression, and in some cases could impair the ability of co-expressed wild-type Mfd to do so. The simplest explanation of these data is that in wild-type cells, Mfd displaces a high proportion of transcription complexes that are stalled by protein roadblocks before the repressor protein dissociates from the DNA. In addition to its role in roadblock repression, Mfd is able to revive backtracked elongation complexes (8), and is responsible for RNAP release during transcription termination by the phage λ Nun protein (34). It is increasingly apparent that Mfd plays a more general role in the regulation of transcription elongation than simply mediating transcription-coupled repair.

Mfd couples ATP hydrolysis to movement of DNA–RNAP complexes, and RecG couples ATP hydrolysis to movement of specific DNA structures. Both proteins therefore require a 'motor' domain. Sequence alignment of the two proteins revealed that, in addition to conserved superfamily II helicase motifs, they share an area of high sequence identity that includes a helical hairpin motif termed the TRG motif (14). As mutational analysis indicated that residues within this motif were important for DNA translocation by RecG (14), it was likely that the TRG motif would also be important for the function of Mfd. We tested this hypothesis by making single amino acid substitutions at the residues within Mfd that correspond to the key residues in the TRG motif of RecG (Fig. 1). All of these substitutions caused defects in Mfd.

Alanine substitutions of R953 and Q963 abolished the ability of Mfd to displace stalled transcription elongation complexes *in vitro*, and produced Mfd derivatives that were unable to complement an *mfd* deletion mutant *in vivo*. These residues therefore appear to play essential roles in the mechanism by which Mfd dissociates RNAP from DNA. Substitution of R929 or H948 impaired the RNAP displacement activity of Mfd *in vitro*, but proteins containing these substitutions retained sufficient activity to complement an *mfd* deletion mutant *in vivo*. The DNA and ATP binding activities of all four mutants appeared to be intact. The 'hierachy' of effects of TRG substitutions in Mfd mirrored that observed in RecG: substitutions of equivalent residues caused defects of equivalent magnitude in both proteins. In Mfd, alanine substitutions of R953 and Q963 caused greater defects than substitution of R929 *in vivo* and *in vitro*. In *E.coli* RecG, substitutions of R630 (equivalent to Mfd R953) and Q640 (equivalent to Mfd Q963) had a greater effect on the *in vivo* role of RecG in UV tolerance than substitutions of R609 (equivalent to Mfd R929), and substitutions of R630 caused greater defects in RecG helicase and ATPase activity *in vitro* than substitutions of R609 (the effects of RecG Q640 substitutions *in vitro* have not been reported) (14). We conclude that the TRG motif of Mfd is essential for Mfd-mediated displacement of stalled RNAP, and that Mfd couples ATP hydrolysis to RNAP displacement via a mechanism very similar to that used by RecG to drive migration of Holliday junctions. Our results therefore support a model for Mfd action in which dissociation of RNAP is dependent on translocation of double-stranded DNA by Mfd.

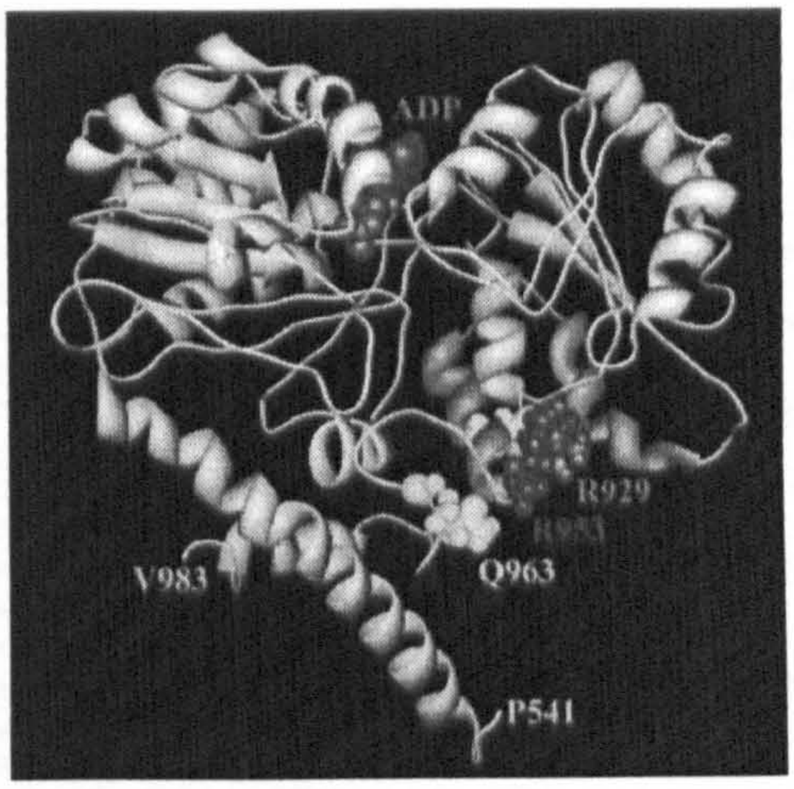


Figure 6. Model of residues 541–983 of Mfd. The model is based on the coordinates of the *T.maritima* RecG-ADP–DNA complex [accession code 1GM5; (12)] and was constructed by homology modelling using ClustalW and InsightII 97 (Accelrys). The image was produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco. The orange ribbon denotes the helical hairpin of the TRG motif, the yellow ribbon denotes the surface-exposed loop of the TRG motif, and the green ribbon denotes helicase motif VI. The residues analysed in this study are labelled, with the exception of H948 (white spheres), which is partially obscured by R953.

The primary sequences of the helicase motifs and TRG motif of Mfd and RecG are highly homologous, and we constructed a model of residues 541–983 of Mfd based on the crystal structure of residues 309–755 of *T.maritima* RecG (Fig. 6). Within this model, Mfd residues R929 and R953 are closely juxtaposed at the base of a helical hairpin, and Mfd residue Q963 falls in a surface-exposed loop that was not resolved in the RecG-ADP–DNA crystal structure. Mfd H948, which is not conserved in RecG or in Mfd sequences from different species, falls within the helical hairpin bounded by R929 and R953, and is adjacent to D949, a highly conserved residue that is implicated in TRG function in RecG (14). R953 is suitably located to form hydrogen bonds with D889 (a residue at the N-terminal end of a helix that extends into helicase motif VI) and with D949.

Mahdi and co-workers (14) suggested a plausible model for the generation of ATP-driven conformational change in the TRG motif, based on the crystal structure of RecG (12). Despite differences in the size of the TRG helical hairpin, and some differences in the residues involved, this model appears to be equally applicable to Mfd. The model, adapted for Mfd, can be summarised as follows: an arginine residue at the C-terminal end of helicase motif VI interacts with the γ phosphate of ATP. The orientation of the helix containing helicase motif VI changes during the cycle of ATP binding, hydrolysis and release, and consequently so does the pattern of hydrogen bonds made between D889 and the arginines within

the TRG motif. This causes a change in the conformation of the helical hairpin structure bounded by R929 and R953, which is transmitted to the loop containing Q963.

As the path of DNA across the translocase region of RecG and Mfd is not well defined, the manner in which this series of conformational changes leads to DNA translocation remains speculative. Mahdi *et al.* (14) have suggested that the conserved glutamine (Q640 in *E.coli* RecG; Q963 in *E.coli* Mfd) might interact directly with DNA, and that the associated surface-exposed loop might drive DNA translocation by acting as a swinging arm. Mfd QA963 retains the ability to bind DNA and it is not currently possible to state whether or not Q963 interacts with DNA during the translocation process. Our structural model of Mfd and the mechanistic model of Mahdi *et al.* (14) offer explanations for the effects of the other Mfd substitutions tested in this work. The positively charged side chains of R929 and R953 are in close proximity, and the removal of electrostatic repulsion that would result from alanine substitution of either of these residues would probably disrupt the fold of the helical hairpin region as well as removing the hydrogen bonding link to ATP via helicase motif VI. The effect of HL948 may be direct, as it results in the removal of an ionisable side chain from the immediate vicinity of R929 and R953, or indirect, as the side chain is immediately adjacent to D949 which forms a hydrogen bond with R953 in the modelled Mfd structure. The identification of the Mfd TRG motif also explains the properties of the deletion mutant Mfd_{trunc}, which was isolated in early studies of Mfd and found to be unable to dissociate RNAP (6): the deletion removes the C-terminal helix of the TRG helical hairpin and the loop containing Q963.

Whilst the ATP-driven motors of Mfd and RecG share a common mechanism, they appear to differ in performance. RecG is an efficient, DNA-stimulated ATPase (k_{cat} >100/s) (13), but Mfd is an inefficient ATPase (k_{cat} ~3/min) whose activity is not enhanced by DNA or by a stalled RNAP complex (5). RecG is able to drive the translocation of Holliday junctions *in vitro* in the absence of other factors (13), but there is as yet no evidence for translocation of naked DNA by Mfd and, unlike some DNA translocating enzymes (10,11), purified Mfd does not displace a triplex-forming oligonucleotide bound to a naked double-stranded DNA template (N.J.Savery, unpublished). The low k_{cat} for ATP hydrolysis may act as a kinetic brake that ensures Mfd competes ineffectively with transcription elongation (8) and avoids displacing transcriptionally competent RNAP complexes, but further investigation is needed to determine whether the ATPase and/or translocase activities of Mfd are stimulated *in vivo* by auxiliary factors.

ACKNOWLEDGEMENTS

We are grateful to Dr R. Sessions for expert assistance with molecular modelling, Professor M. J. Chamberlin and Dr C. Selby for providing materials, Professor R. G. Lloyd for communicating results prior to publication, Professor S. E. Halford and Dr M. D. Szczelkun for discussions, and Ms C. S. Bell for technical assistance. This work was supported by the Biotechnology and Biological Sciences Research Council via a studentship for A.C. and research grant G15249.

REFERENCES

1. Mellon,I. and Hanawalt,P.C. (1989) Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature*, **342**, 95–98.
2. Mellon,I., Spivak,G. and Hanawalt,P.C. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*, **51**, 241–249.
3. Svejstrup,J.Q. (2002) Mechanisms of transcription-coupled DNA repair. *Nature Rev. Mol. Cell. Biol.*, **3**, 21–29.
4. Selby,C.P. and Sancar,A. (1993) Transcription–repair coupling and mutation frequency decline. *J. Bacteriol.*, **175**, 7509–7514.
5. Selby,C.P. and Sancar,A. (1993) Molecular mechanism of transcription–repair coupling. *Science*, **260**, 53–58.
6. Selby,C.P. and Sancar,A. (1995) Structure and function of transcription–repair coupling factor. I. Structural domains and binding properties. *J. Biol. Chem.*, **270**, 4882–4889.
7. Selby,C.P. and Sancar,A. (1995) Structure and function of transcription–repair coupling factor. II. Catalytic properties. *J. Biol. Chem.*, **270**, 4890–4895.
8. Park,J.S., Marr,M.T. and Roberts,J.W. (2002) *E.coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell*, **109**, 757–767.
9. Borukhov,S., Sagitov,V. and Goldfarb,A. (1993) Transcript cleavage factors from *E.coli*. *Cell*, **72**, 459–466.
10. Firman,K. and Szczelkun,M.D. (2000) Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO J.*, **19**, 2094–2102.
11. Whitehouse,I., Stockdale,C., Flaus,A., Szczelkun,M.D. and Owen-Hughes,T. (2003) Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Mol. Cell. Biol.*, **23**, 1935–1945.
12. Singleton,M.R., Scaife,S. and Wigley,D.B. (2001) Structural analysis of DNA replication fork reversal by RecG. *Cell*, **107**, 79–89.
13. Lloyd,R.G. and Sharples,G.J. (1993) Dissociation of synthetic Holliday junctions by *E.coli* RecG protein. *EMBO J.*, **12**, 17–22.
14. Mahdi,A.A., Briggs,G.S., Sharples,G.J., Wen,Q. and Lloyd,R.G. (2003) A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. *EMBO J.*, **22**, 724–734.
15. Sambrook,J. and Russel,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16. Burns,H. and Minchin,S. (1994) Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. *Nucleic Acids Res.*, **22**, 3840–3845.
17. Simons,A., Tils,D., von Wilcken-Bergmann,B. and Muller-Hill,B. (1984) Possible ideal *lac* operator: *Escherichia coli lac* operator-like sequences from eukaryotic genomes lack the central G X C pair. *Proc. Natl Acad. Sci. USA*, **81**, 1624–1628.
18. Lodge,J., Fear,J., Busby,S., Gunasekaran,P. and Kamini,N.R. (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol. Lett.*, **74**, 271–276.
19. Stanley,L.K. and Savery,N.J. (2003) Characterisation of the *Escherichia coli mfd* promoter. *Arch. Microbiol.*, **179**, 381–385.
20. Zhou,Y.H., Zhang,X.P. and Ebright,R.H. (1991) Random mutagenesis of gene-sized DNA molecules by use of PCR with *Taq* DNA polymerase. *Nucleic Acids Res.*, **19**, 6052.
21. Savery,N.J., Lloyd,G.S., Kainz,M., Gaal,T., Ross,W., Ebright,R.H., Gourse,R.L. and Busby,S.J.W. (1998) Transcription activation at class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase α subunit. *EMBO J.*, **17**, 3439–3447.
22. Levin,J.R., Krummel,B. and Chamberlin,M.J. (1987) Isolation and properties of transcribing ternary complexes of *Escherichia coli* RNA polymerase positioned at a single template base. *J. Mol. Biol.*, **196**, 85–100.
23. Fried,M.G. (1989) Measurement of protein–DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis*, **10**, 366–376.
24. Tombline,G. and Fishel,R. (2002) Biochemical characterization of the human RAD51 protein. I. ATP hydrolysis. *J. Biol. Chem.*, **277**, 14417–14425.
25. Miwa,Y. and Fujita,Y. (1993) Promoter-independent catabolite repression of the *Bacillus subtilis gnt* operon. *J. Biochem.*, **113**, 665–671.

26. Wray,L.V.,Jr and Fisher,S.H. (1994) Analysis of *Bacillus subtilis hut* operon expression indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms: regulation of transcription initiation and inhibition of histidine transport. *J. Bacteriol.*, **176**, 5466–5473.
27. Zalieckas,J.M., Wray,L.V.,Jr, Ferson,A.E. and Fisher,S.H. (1998) Transcription–repair coupling factor is involved in carbon catabolite repression of the *Bacillus subtilis hut* and *gnt* operons. *Mol. Microbiol.*, **27**, 1031–1038.
28. Abo,T., Inada,T., Ogawa,K. and Aiba,H. (2000) SsrA-mediated tagging and proteolysis of LacI and its role in the regulation of *lac* operon. *EMBO J.*, **19**, 3762–3769.
29. Deuschle,U., Gentz,R. and Bujard,H. (1986) *lac* repressor blocks transcribing RNA polymerase and terminates transcription. *Proc. Natl Acad. Sci. USA*, **83**, 4134–4137.
30. Sellitti,M.A., Pavco,P.A. and Steege,D.A. (1987) *lac* repressor blocks *in vivo* transcription of *lac* control region DNA. *Proc. Natl Acad. Sci. USA*, **84**, 3199–3203.
31. Guerin,M., Leng,M. and Rahmouni,A.R. (1996) High resolution mapping of *E.coli* transcription elongation complex *in situ* reveals protein interactions with the non-transcribed strand. *EMBO J.*, **15**, 5397–5407.
32. Muller,J., Oehler,S. and Muller-Hill,B. (1996) Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J. Mol. Biol.*, **257**, 21–29.
33. Krummel,B. and Chamberlin,M.J. (1992) Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Individual complexes halted along different transcription units have distinct and unexpected biochemical properties. *J. Mol. Biol.*, **225**, 221–237.
34. Washburn,R.S., Wang,Y. and Gottesman,M.E. (2003) Role of *E.coli* transcription–repair coupling factor Mfd in Nun-mediated transcription termination. *J. Mol. Biol.*, **329**, 655–662.